GENETIC INTERRELATIONS OF TWO ANDROMONOECIOUS TYPES OF MAIZE, DWARF AND ANTHER EAR[1]

R. A. EMERSON AND STERLING H. EMERSON

Cornell University, Ithaca, New York

Received July 14, 1921

TABLE OF CONTENTS

INTRODUCTION 203
Characteristics of the dwarf and anther-ear types 204
Inheritance of the dwarf type 210
Inheritance of the anther-ear type 212
Genetic interrelations of the dwarf and anther-ear types 212
Variability in stature of dwarf and anther-ear types 218
Rudimentary stamens in dwarf and anther-ear types 224
SUMMARY 225
LITERATURE CITED 226
APPENDIX—TABLES 227

INTRODUCTION

Attention was called by MONTGOMERY (1906) to the occasional appearance of perfect flowers in the staminate inflorescence of maize and similar cases were reported by KEMPTON (1913). MONTGOMERY (1911) described with illustrations a true-breeding type of semi-dwarf dent maize, the ears of which were perfect-flowered. Perfect-flowered maize was described and illustrated also by BLARINGHEM (1908, pp. 180-183). EAST and HAYES (1911, pp. 13, 14) noted and illustrated a perfect-flowered sweet corn. WEATHERWAX (1916, 1917) showed that typically pistillate flowers of maize exhibit in microscopic sections the rudiments of stamens and that staminate flowers show rudiments of pistils.

The senior author of this paper (EMERSON 1911, pp. 82-87; 1912) described and illustrated dwarf and semi-dwarf andromonoecious maize types and showed that they are inherited as simple recessives to monoecious forms of normal height. Later work has shown that two genetically distinct types, now known as "dwarf" and "anther ear," were confused in these accounts. It was noted in one of them (EMERSON 1911) that the andromonoecious forms of unrelated stocks differed much in height, but

[1] Paper No. 90, Department of Plant Breeding, CORNELL UNIVERSITY, Ithaca, New York.

the differences were thought to be due to growth factors modifying the height of dwarfs as multiple factors are assumed to differentiate normal races of diverse heights.

CHARACTERISTICS OF THE DWARF AND ANTHER-EAR TYPES

Perhaps the most nearly unique feature of both dwarf and anther ear is the presence of stamens throughout their ears (figures 1 and 2). Sometimes one flower of the spikelet has three well developed stamens and the other flower more or less rudimentary ones but in other cases both flowers have fully developed stamens. Since the terminal inflorescences of dwarf and anther-ear plants have staminate flowers only, the types are andromonoecious. Exceptions to this condition, that is, dwarf and anther-ear plants without ear stamens or with rudimentary ones only, have been observed and are discussed in a later section of this account.

The ears of both dwarf and anther ear generally end in thick unbranched tassel-like spikes having staminate flowers only. Anther-ear plants in which the ear is represented by only a few seeds at the base of these club-like staminate-flowered spikes are common, and one plant was found that lacked even these few seeds and that was therefore wholly staminate-flowered. Some stocks of both dwarf and anther ear, on the contrary, have no tassel-like appendages to their ears. In any case the ears are short and relatively thick (figure 1).

The tassels of dwarf plants are compact and little branched, often quite unbranched. The smaller forms of the anther-ear type have tassels that are very similar to those of dwarfs in both compactness and branching while the taller forms usually have tassels that are more slender and more branched than those of dwarfs but rarely so much so as those of normal plants (figures 3 and 4). Our early cultures of dwarfs shed little or no pollen, due apparently to failure of the glumes to open normally. Pollen collected from anthers that had been removed from the tassel proved to be functional. Two dwarf plants found in a culture grown in 1913 shed pollen abundantly. By selection in later cultures of dwarfs descended mostly from these two plants, dwarf strains have been obtained that shed pollen fairly well.

The dwarf type is further characterized by very short culms, ordinarily not more than one-fourth the length of the culms of related normal plants (figure 3). Anther-ear plants also are usually considerably shorter than normals. There are, however, marked differences between different stocks of anther ear with respect to height, some being definitely semi-dwarf and some only slightly shorter than related normal plants (figures 3 and 4).

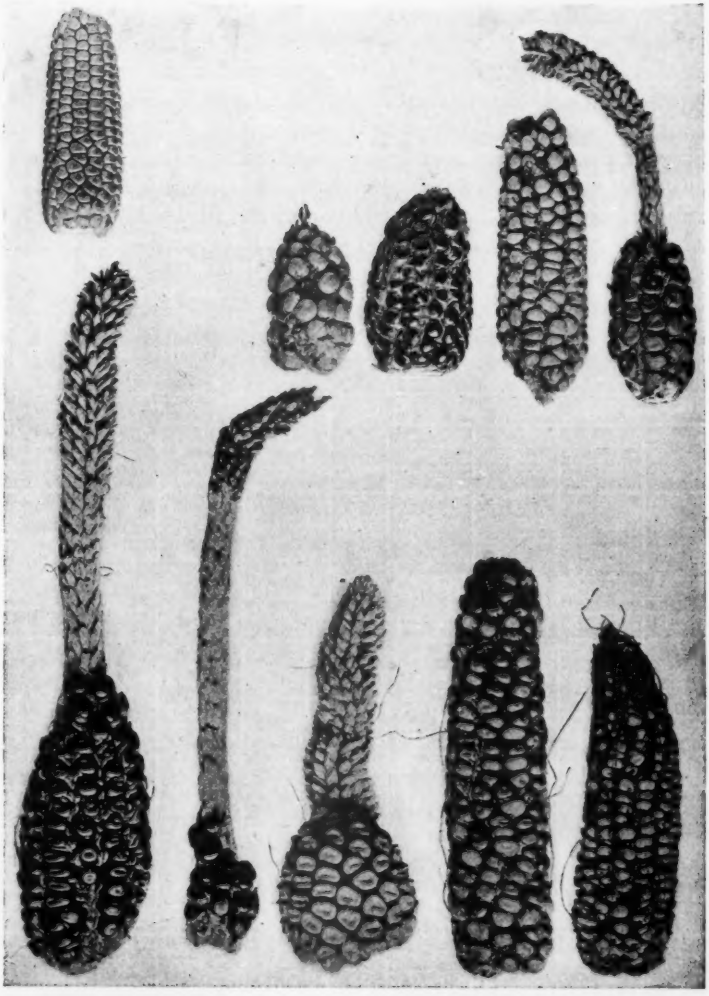

FIGURE 1.—Representative ear types. Below, five anther ear; above, right, four dwarf; above, left, one Tom Thumb.

The possible significance of this variation in height is considered in a later section of this paper. The differences in height are apparent even in the seedling stage (figure 5), though the taller forms of anther-ear seedlings are not always greatly different from normals.

The leaves of dwarf plants are characteristically short and broad. The first seedling leaves are not only shorter but actually broader than those of normal plants and the leaves of mature dwarf plants, though very short, are often quite as broad as those of normal plants. The leaves of most anther-ear plants also are relatively short and broad, being in this respect as a rule intermediate between those of dwarf and those of normal plants. The leaves of the taller anther-ear stocks are often not very different from those of normal plants.

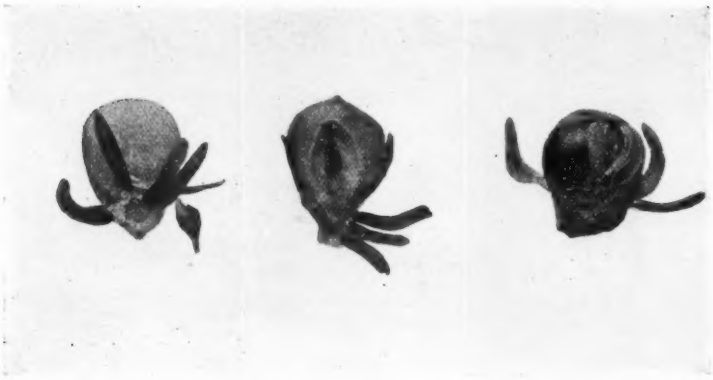

FIGURE 2—Seeds of anther ear and dwarf with stamens attached.

The low stature of dwarf and of anther-ear plants is due to shorter internodes than those of normal plants. In number of internodes these types are not characteristically different. In contrast to this condition of dwarf and anther ear, the low stature of such varieties as Tom Thumb pop is due to the small number of internodes, not merely to short ones (figure 4). Tom Thumb is normal not only in lacking ear stamens (figure 1) but in the relative length, breadth and thickness of its parts. Its ears, tassels, culms and leaves are like those of normal tall plants except that they are proportionately smaller in all dimensions. Dwarf and anther ear, on the other hand, are dwarfed in all longitudinal dimensions but not correspondingly so in diameter of parts.

A type of extreme dwarfness, known as "dwarf anther ear," the double recessive of dwarf and anther ear, as its name indicates, has appeared in certain crosses discussed in a later part of this account (figures 3, 5 and 6). This type is much smaller even than the type called "dwarf." It is char-

FIGURE 3.—Mature plants of normal, anther ear, dwarf, and dwarf anther ear.

acterized by an excessive shortening of all longitudinal dimensions. The length of culm, including the tassel, is often not over 12 centimeters. Its largest leaves are not far from 12 centimeters long and 6 centimeters

broad. Since the plant usually has several tillers which spread out almost on the ground, it sometimes has a total spread of branches and leaves of

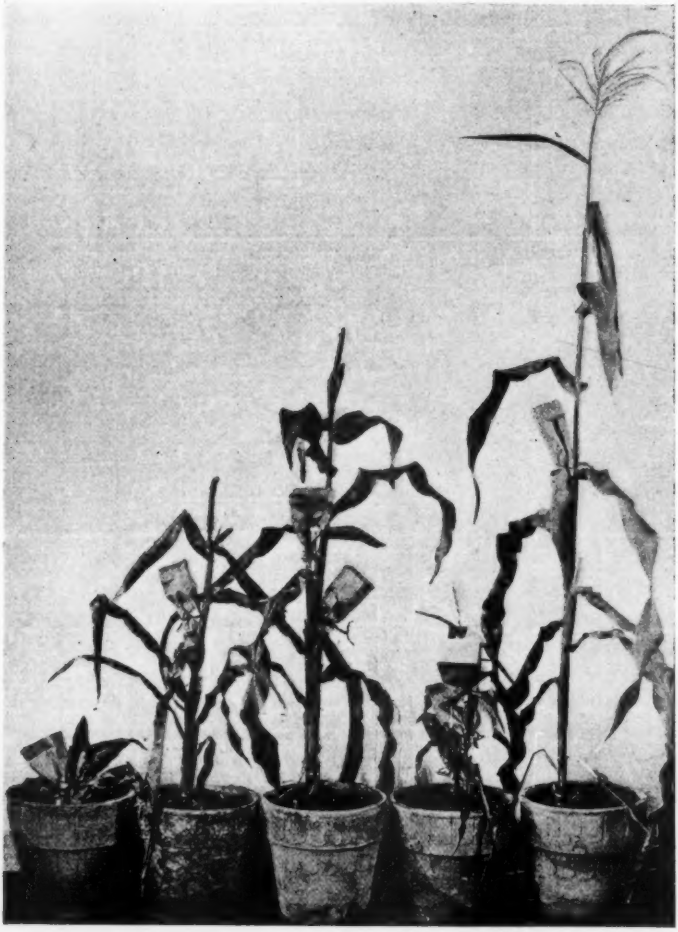

FIGURE 4.—Mature plants of dwarf, small anther ear, larger anther ear, very small normal Tom Thumb, and tall normal.

two and a half times its height. The tassel is a short, compact, unbranched spike. The anthers do not shed pollen normally. No adequate test has

FIGURE 5.—Seedlings of normal, anther ear, dwarf, and dwarf anther ear.

been made to determine whether or not the pollen is functional. No fertile ear shoots have been observed on any of these plants.

INHERITANCE OF THE DWARF TYPE

That the dwarf type is inherited as a simple recessive to normal has been reported previously by the senior author (EMERSON 1911, 1912). Since in these accounts the anther-ear type was confused with the dwarf type, it seems wise to present here such evidence as is necessary to show the mode of inheritance of the true dwarf type in crosses between dwarf and normal.

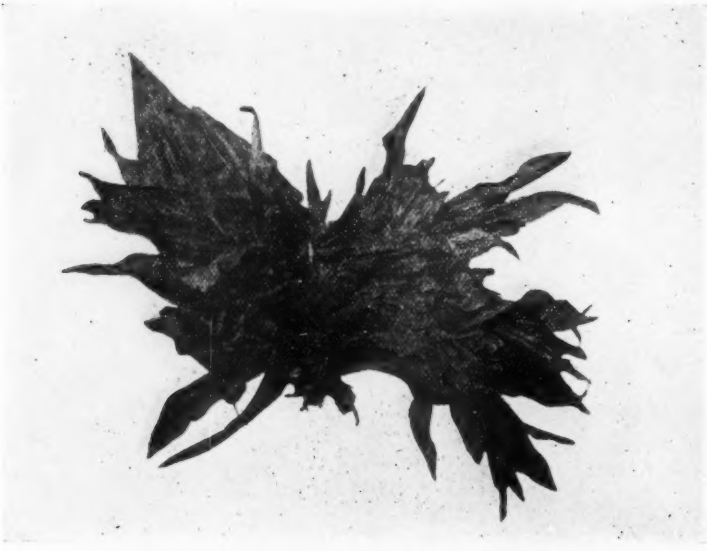

FIGURE 6.—Mature plant of dwarf anther ear.

Our records show that nineteen direct crosses of dwarf with normal resulted in 272 F_1 plants, all of which were normal. From selfed normal plants heterozygous for dwarf, fifty-four F_2 progenies with a total of 2932 individuals were grown. Of these F_2's thirty progenies were grown from seed planted in the field (table 7, group 1; see appendix). They resulted in 1095 normal and 231 dwarf plants. The deviation from a 3 : 1 ratio is 100.5 ± 10.6. The odds against such a deviation being due to chance are almost inconceivably great. The longitudinal axis of germinating seedlings of dwarf is relatively as short as that of mature plants.

Part of these F_2 cultures were grown at Lincoln, Nebraska, where it is customary to plant maize seeds about ten centimeters deep. The other cultures were grown at Ithaca, N. Y., from plantings not over three or four centimeters deep but in sticky clay soil. It seems obvious that under neither of these conditions could dwarf be expected to germinate well.

Fourteen other F_2 cultures (table 7, group 2) were germinated in light soil in greenhouse flats from seed planted about two centimeters deep. These cultures contained 436 normals and 138 dwarfs, a deviation from the 3 : 1 ratio of 5.5 ± 7.0. These two lots of F_2's were closely related, all having come from a single lot a few generations before, so that there can be little doubt that the same type is concerned in both. More decisive evidence, however, is afforded by the behavior of ten other F_2 cultures (table 7, group 3), each of which was planted both in the field and in the greenhouse. Of the field-grown lots, 277 were normal and 60 dwarf, a deviation from the 3 : 1 ratio of 24.3 ± 5.4, which should not occur by chance more than once in somewhat over 400 trials. Of the lots germinated in the greenhouse, there were 510 normals and 185 dwarfs, an excess of dwarfs over expectation of 11.3 ± 7.7, or such a deviation as might occur by chance about once in three trials.

F_1 plants of this cross, back-crossed with dwarfs, yielded eight progenies (table 8) with 331 normals and 354 dwarfs, a deviation from the expected equality of 11.5 ± 8.8, or such as might occur by chance almost twice in five trials. All but one of these eight progenies were germinated in the greenhouse.

Some of the F_2 normals remained constant in F_3 and others again threw dwarfs. Normals of the segregating F_3 progenies should theoretically be equivalent to F_2's. In all, seven F_3 and equivalent F_4 progenies (table 9, group 1) gave none but normal plants, a total of 208. Fifteen F_3 and F_4 progenies (table 9, group 2) gave 484 normals and 120 dwarfs, a deviation from the 3 : 1 relation of 31.0 ± 7.2, or such as might occur by chance only once in about 270 trials. All but three of the fifteen progenies were grown in the field. It is expected that one in three F_2 and equivalent F_3 normal plants will breed true. The seven true-breeding and fifteen segregating progenies is a close approach to this expectation.

Five F_2 and F_3 dwarfs bred true in the next generation (table 9, group 3) having produced a total of 56 dwarfs and one normal, the latter doubtless due to accidental pollination.

It is to be concluded, therefore, that dwarf, *d d*, behaves as a simple Mendelian recessive to normal, *D D*.

INHERITANCE OF THE ANTHER-EAR TYPE

That anther ear, like dwarf, is a simple Mendelian recessive to normal is shown by the data now to be presented. Eighteen F_1 progenies of crosses between anther ear and normal resulted in 485 normal plants. From such heterozygous normal plants, twenty-seven F_2 progenies with a total of 1151 individuals were grown. Of these, twelve progenies grown in the field (table 10, group 1) had 455 normal and 108 anther-ear plants, a deviation from a 3 : 1 ratio of 32.8 ± 6.9, or such as would be expected to occur by chance not more than once in about 750 trials. Since anther ear is less vigorous than normal, it seems probable that this deficiency of anther ear was due to the failure of some plants to survive under field conditions. Thirteen F_2 progenies germinated in the greenhouse (table 10, group 2) contained 320 normal and 128 anther-ear plants, a deviation of 16.0 ± 6.2 from a 3 : 1 ratio. Such a deviation might be expected to occur by chance once in about twelve such trials. Two progenies (table 10, group 3) were grown in both the field and greenhouse. Of the field-grown plants there were 53 normal and 8 anther ear and of the greenhouse-grown ones 55 normal and 24 anther ear. The deviations from a 3 : 1 relation are 7.3 ± 2.3 and 4.3 ± 2.6, or such as might occur by chance about once in thirty and once in four trials, respectively.

Four normal plants of F_2 bred true in F_3 (table 11, group 1), three of them having produced a total of 272 normal plants from self-pollination and one of them 12 normals from a back-cross with anther ear. Seven F_2 normals and one F_3 equivalent to an F_2 (table 11, group 2) produced 393 normal and 110 anther-ear plants, a deviation from the 3 : 1 ratio of 15.8 ± 6.6, or such as might occur by chance about once in nine trials. The deficiency of anther ear is probably due to the fact that all but one of these cultures were field-grown.

Three F_2 anther-ear plants (table 11, group 3) produced 36 anther-ear and 2 normal plants in F_3. The normals are doubtless to be accounted for by accidental pollination, so that the conclusion that anther-ear plants breed true seems justifiable.

On the whole, therefore, anther ear, a_n a_n, can be said to behave in inheritance as a simple recessive to normal, A_n A_n.

GENETIC INTERRELATIONS OF THE DWARF AND ANTHER-EAR TYPES

In the foregoing account, it has been assumed that dwarf and anther ear are genetically different types. The evidence upon which this assumption is based is presented here. It has been stated previously that the two

types were confused in the earlier papers of the senior author. Both types exhibit so much variation in stature and have so many characters in common,—stamens in the ears, relatively short culms, leaves, tassels and ears, and recessiveness to normal,—that this confusion was perhaps not unnatural in the early stages of the work.

Crosses of dwarf, *d*, with anther ear, a_n, have resulted in plants that were normal in all respects, including relatively long culms and leaves, slender much-branched tassels, and ears without stamens or with mere rudiments of stamens. Six such crosses (table 12) gave F_1 progenies totaling 83 normal plants. In addition normal plants heterozygous for dwarf, crossed with normals heterozygous for anther ear, resulted in 21 normal plants.

The F_2 progenies of the cross dwarf × anther ear, $D\ d\ A_n\ a_n$, are recorded in group 1 of table 13. A cross of a normal, $D\ D\ A_n\ A_n$, with an anther ear that is heterozygous for dwarf, $D\ d\ a_n\ a_n$, should give two F_1 genotypes, $D\ D\ A_n\ a_n$ and $D\ d\ A_n\ a_n$. Similarly a cross of a normal with a dwarf that is heterozygous for anther ear, $d\ d\ A_n\ a_n$, should give two genotypes, $D\ d\ A_n\ A_n$ and $D\ d\ A_n\ a_n$. The double heterozygotes, in case of each of these crosses, are equivalent to F_1's of the cross normal × dwarf anther ear and should give the same results in F_2 as F_1's of the cross dwarf × anther ear. Two such F_2 progenies are recorded in group 2 of table 13, the first being a cross of normal with anther ear and the second a cross of normal with dwarf.

The eight F_2 progenies had a total of 752 individuals distributed as shown below to the four types, normal, dwarf, anther ear, and dwarf anther ear. There is about one chance in thirteen that the deviations from the 9 : 3 : 3 : 1 relation, expected on the basis of two independent factor pairs, may be due to errors of random sampling, P equaling 0.076. The comparison follows:

	Normal	*Dwarf*	*Anther ear*	*Dwarf anther ear*	*Total*
Observed.....	446	139	136	31	752
Calculated.....	*423*	*141*	*141*	*47*	*752*
Difference.....	+23	−2	−5	−16	0

The deficiency of dwarf anther ear is doubtless due to the extremely small size of this type, which makes it incapable of germinating or of surviving under field conditions. Since both dwarf and anther ear are ordinarily deficient in field cultures, the much smaller double recessive could hardly be expected to survive in the field. The apparent excess of normals is probably due to the deficiencies of the other three types. That this explanation of the rather poor fit of the observed with the theoretical distributions is correct is made highly probable by the comparisons given

below. Of the eight F_2 progenies, two were grown in the field only, three in the greenhouse only, and three in both the field and the greenhouse. Of the field-grown lots the deviations from the calculated frequencies are so great that they would not be expected to occur by chance more than about once in over twelve thousand trials, P equaling 0.000,085. With the greenhouse-grown lots, on the other hand, there is so close a fit between the observed and the theoretical frequency distributions that no value of P is to be found in published probability tables, χ^2 equaling 0.65. (When $\chi^2=1$ and $n'=4$, $P=0.80$.) The comparisons follow:

	Normal	*Dwarf*	*Anther ear*	*Dwarf anther ear*	*Total*
Field cultures					
Observed.....	182	43	49	1	275
Calculated.....	*155*	*52*	*52*	*17*	*276*
Difference.....	+27	−9	−3	−16	−1
Greenhouse cultures					
Observed.....	264	96	87	30	477
Calculated.....	*268*	*89*	*89*	*30*	*476*
Difference.....	−4	+7	−2	0	+1

Three F_1 normals back-crossed with anther ear gave 157 normals and 93 anther ears (table 14, group 1). Numerical equality of these two classes is to be expected on the basis of independent inheritance of the complementary factor pairs *D d* and A_n a_n. The deviation from expectation is 32.0 ± 5.3. This extreme deficiency of anther ear must be explained by the fact that all these cultures were field-grown, mostly under unfavorable conditions.

Five F_1 normals of the cross of dwarf with anther ear were back-crossed with dwarf and gave 274 normal and 204 dwarf plants (table 14, group 2). The deviation from equality is 35.0 ± 7.4, which, while less than that found in the case of anther ear from back-crosses of F_1 with anther ear, is so great that appeal must be made here also to the difficulties to which dwarf is subjected under unfavorable field conditions.

The F_2 results of the cross dwarf × anther ear reported above, particularly the greenhouse-grown cultures, leave little doubt that *D d* and A_n a_n are independent complementary factor pairs. The crucial evidence, however, must be sought in the F_3 behavior of the several F_2 types. On the basis of the hypothesis under test, the F_2 types should behave as follows:

F_2 types		F_3 behavior			
Phenotype	Genotype	Normal	Dwarf	Anther ear	Dwarf anther ear
9 normal	1 *D D* A_n A_n	1			
	2 *D d* A_n A_n	3	1		
	2 *D D* A_n a_n	3		1	
	4 *D d* A_n a_n	9	3	3	1
3 dwarf	1 *d d* A_n A_n		1		
	2 *d d* A_n a_n		3		1
3 anther ear	1 *D D* a_n a_n			1	
	2 *D d* a_n a_n			3	1
1 dwarf anther ear	1 *d d* a_n a_n				1

Twenty-four F_2 normals were tested in F_3 (table 15). Plantings of most of the F_3 progenies (culture numbers S48-S90) were made in the greenhouse, and duplicate plantings were made in the field under unusually adverse soil conditions, part of them (S126-S166) in 1918 and part (9618-9642) in 1919. It was hardly to be expected that the weaker types would appear in field plantings in numbers close to the theoretical ones. The field and greenhouse plantings are recorded together in table 15.

Of the twenty-four F_2 normals tested, two (table 15, group 1) remained constant, *D D* A_n A_n, in F_3, having produced 290 normal plants and no other types.

Five F_2 normals, *D d* A_n A_n, (group 2) produced 374 normals and 86 dwarfs, a deviation from the 3 : 1 relation of 29.0 ± 6.3. Such a deviation could occur by chance not more than once in over five hundred trials. Of these 460 F_3's, the 173 grown in the greenhouse consisted of 120 normals and 53 dwarfs, an excess over expectation of 9.8 ± 3.8 dwarfs, a deviation that might occur by chance about once in thirteen trials.

Three F_2 normals, *D D* A_n a_n, (group 3) produced 249 normals and 61 anther ears in F_3, a deviation from the 3 : 1 relation of 16.5 ± 5.1, or such as might occur by chance not more than three times in one hundred trials. Of these 310 F_3's, the 132 greenhouse-grown plants consisted of 98 normal and 34 anther-ear plants, a deviation of only 1.0 ± 3.4 from the 3 : 1 ratio.

Fourteen F_2 normals, *D d* A_n a_n, (table 15, group 4) produced in F_3 the three types, normal, dwarf and anther ear, and six of them produced dwarf anther ear in addition. In a total of 1174 F_3 plants, the four types, in the order given above, appeared in the numbers 774 : 174 : 215 : 11.

The departures from expectation are so extreme that no value of P is to be found in published tables, χ^2 equaling 84. (When $\chi^2 = 30$ and $n' = 4$, $P = 0.000{,}001$). Of these 1174 F_3 plants, 594 that were germinated in the greenhouse exhibited the four types in the numbers 330 : 105 : 148 : 11. Even in case of these greenhouse-grown seedlings, there is not more than one chance in a million that the observed deviations from expectation could be due to errors of random sampling. The great deficiency in the very weak double recessive, dwarf anther ear, is responsible more than anything else for the poor fit. All the plants of this type that appeared were in greenhouse cultures but apparently they were too weak to complete their germination even under greenhouse conditions. The only other marked difference between observed and calculated frequencies in these greenhouse cultures occurred in the anther-ear class, but here there was an excess over expectation. It is not particularly difficult to separate the smaller types of anther ear from normal even in the seedling stage, but the taller anther-ear stocks cannot always be distinguished until the flowering stage. It seems possible that the recorded excess of anther ear in these lots may have come in part from errors in classification.

Notwithstanding the poor fit of observed with calculated frequencies in some cases, the F_2 normals of the cross of dwarf with anther ear exhibited in general the F_3 behavior expected of them in accordance with the hypothesis that *D d* and $A_n\ a_n$ are complementary independently inherited factor pairs. Of the twenty-four F_2 normals tested, the four types of F_3 behavior, were exhibited in the relation of 2 : 5 : 3 : 14, while the theoretical relation for twenty-four individuals is 2.7 : 5.3 : 5.3 : 10.7. There is slightly better than an even chance that deviations such as this might occur through errors of random sampling.

Two F_2 dwarfs were tested by their F_3 behavior, both in greenhouse plantings. One of these, $d\ d\ A_n\ A_n$, (table 15, group 5) bred true, having produced 51 dwarf plants, and one, $d\ d\ A_n\ a_n$, threw 17 dwarfs and 6 dwarf anther ears, as nearly a 3 : 1 ratio as is possible with a total of 23 plants. Theoretically, one out of three F_2 dwarfs should breed true and two throw dwarfs and dwarf anther ears.

No F_2 anther ear was tested in F_3. It is obviously impossible to test dwarf anther ear, since no plant of that type has ever produced seeds.

It has been shown (table 14) that F_1 normals of the cross dwarf × anther ear back-crossed with anther ear throw only normal and anther ear in the next generation. Of such normals, half, $D\ D\ A_n\ a_n$, should again throw normal and anther ear, and half, $D\ d\ A_n\ a_n$, should throw all four types. Records of twenty-one such tests are given in table 16. Eleven normals

threw normal and anther ear (table 16, group 1) in the relation of 252 : 63, a deviation from 3 : 1 of 15.8 ± 5.2. Only one of these progenies was germinated in the greenhouse and this consisted of 73 normal and 24 anther ear, deviation 0.3 ± 2.9. The other ten normals tested (group 2) gave a total of 257 plants distributed to the four types in the relation of 186 : 26 : 43 : 2. This extreme departure from the expected 9 : 3 : 3 : 1 relation is doubtless due to the fact that all but one of the cultures were field-grown. This one showed the four types in the relation of 10 : 3 : 3 : 2, a very close fit to expectation, χ^2 equaling 0.83. Of the 21 normals tested, 11 gave the one type of behavior and 10 the other, where equality was expected.

Nine anther-ear plants of these same back-crosses (table 8) were tested by their F_3 behavior. Four bred true, *D D* a_n a_n, with a total of 57 anther-ear plants (table 16, group 3) and five, *D d* a_n a_n, produced anther ear and dwarf anther ear only (group 4) in the relation of 127 : 37, a deviation from 3 : 1 of 4.0 ± 3.7. All these lots were grown in the greenhouse. Since half of the anther-ear plants tested are expected theoretically to breed true and half to throw the two types observed, the results are in the closest possible accord with expectation.

F_1 normal plants of the cross dwarf × anther ear back-crossed with dwarf (table 14), it will be recalled, threw normals and dwarfs only. Thirteen of these normals were tested by their progenies. Half of such normals, *D d* A_n A_n, should throw normal and dwarf, and half, *D d* A_n a_n, should throw the four types. Seven of the thirteen (table 17, group 1) threw a total of 161 normal and 26 dwarf, a deviation from 3 : 1 of 20.8 ± 4.0. The deficiency of dwarfs was doubtless due to the fact that all the cultures were field-grown. The other six normals that were tested (table 17, group 2) threw normal, dwarf, and anther ear in the relation of 115 : 10 : 19. The failure of dwarf anther ear to appear, as well as the extreme departure from the expected 9 : 3 : 3 : 1 relation, is ascribed to the field conditions under which all of these cultures were grown.

Three dwarfs from the same back-crosses (table 14) were tested by greenhouse progenies and found (table 17, group 3) to breed true, having produced a total of 173 dwarf plants. Half of them, *d d* A_n A_n, were expected to breed true and half, *d d* A_n a_n, to throw dwarf and dwarf anther-ear. While neither of these dwarfs threw dwarf anther-ear, it has been shown that some such dwarfs carry the a_n factor in a heterozygous condition, *d d* A_n a_n. Dwarfs of back-cross progenies when crossed with normals gave normals in F_1. One of these F_1 normals (table 13, group 2) threw

the four types in F_2 with frequencies of 47 : 12 : 13 : 4. This behavior could have resulted only from a dwarf of the genotype $d\ d\ A_n\ a_n$.

By way of summary, it may be noted that all the genotypes theoretically possible on the assumption of two independently inherited factor pairs, $D\ d$ and $A_n\ a_n$, have been demonstrated in the cross of dwarf with anther ear. It is true that the observed frequency distributions of the several phenotypes have in many instances been far from the theoretical frequencies, but it has been shown conclusively, by growing cultures in both the field and the greenhouse, that these departures from expectation have been due to the failure of the weaker forms to germinate or to survive under field conditions. In a few cases, an excess of anther ear over expectation has been recorded with cultures germinated in the greenhouse, but this result is regarded as due mainly to the difficulty of separating anther ear from normal in the seedling stage.

VARIABILITY IN STATURE OF DWARF AND ANTHER-EAR TYPES

In the foregoing sections of this paper, repeated references have been made to the differences in height between different stocks of anther ear. Similar differences exist between diverse stocks of dwarf. Since all our stocks of dwarf and of anther ear came each from a single plant, it is of interest to consider the possible nature of these diversities.

The dwarf type ordinarily varies in height from about 30 to 60 centimeters as contrasted with heights of from about 120 to 200 centimeters for normal plants of the same cultures. Frequency distributions for height of plants of F_2 progenies of crosses of dwarfs with normals are given in table 1. In no case have the heights of dwarfs and of normals of the

TABLE 1

Heights of normal and dwarf plants of F_2 progenies of the cross dwarf × normal.

PEDIGREE CULTURE	HEIGHT OF PLANTS IN DECIMETERS																		
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Mean
396 Normal....										1	1	3	5	2	2	5	2		16.0
Dwarf.....	1	1																	3.5
7424 Normal....										2	3	7	5	6	7	4		2	15.6
Dwarf.....	11	23	12	3															4.1

same culture been observed to overlap. This sharp segregation in height is ascribed to the single factor pair $D\ d$. That other growth factors, and environmental influences as well, affect the height of dwarfs, much as they do the height of normal plants, is evident from crosses of dwarfs, $d\ d$, with normals $D\ D$, of very low stature, such as Tom Thumb pop.

It has been shown (EMERSON and EAST 1913) that, when a tall normal strain is crossed with Tom Thumb, the height of the F_1 plants is intermediate between the height of the parents, while the range in height of the F_2 plants extends practically to the extremes of both parents. An illustration of this behavior is given in table 2. Such behavior in F_2 is ordinarily interpreted on the basis of partially dominant multiple factors affecting growth. Obviously no single factor pair is alone concerned here, of the nature of *D d*, which is assumed to be responsible for the sharp segregation shown in table 1.

TABLE 2

Heights of parental, F_1, and F_2 plants of the cross tall dent × Tom Thumb pop.

PEDIGREE CULTURE	HEIGHT OF PLANTS IN DECIMETERS																		
	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Mean
357 Tall dent...											2	1	2	6	2	4	3	3	20.9
508 Tom Thumb	2	6	5	2															8.5
509 F_1.....									1	3	5	7	4						17.5
510 F_2.....	2			2	8	6	25	26	29	30	32	28	22	6	4	1	2		15.9

When dwarf plants, *d d*, are crossed with normals, *D D*, of very low stature like Tom Thumb pop, both the normals and dwarfs differentiated by *D d* should be influenced by diverse partially dominant growth factors just as is true of the F_2 plants of crosses of tall normals with the small Tom Thumb. The results of such a cross are presented in table 3. For ready comparison, the dwarfs of culture 7424 (table 1) and Tom Thumb plants of culture 508 (table 2) are included. Neither of these cultures was grown during the same season as the F_1 and F_2 progenies of the cross of dwarf with Tom Thumb. It will be noted that the F_1 plants of this cross were taller than either parent. Evidently the dwarf parent of the cross brought into the combination partially dominant growth factors that

TABLE 3

Heights of Tom Thumb and dwarf plants and of F_1 and F_2 of the cross dwarf × Tom Thumb.

PEDIGREE CULTURE	HEIGHT OF PLANTS IN DECIMETERS																	Mean
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
508 Tom Thumb.							2	6	5	2								8.5
7424 Dwarf.....			11	23	12	13												4.1
10727 F_1 normal....											1	2		4	2	1	1	14.0
10728 F_2 normal....						1	4	3	4	4	6	14	13	12	2	4		11.6
F_2 dwarf.....	1	8	6															2.3

were not present in Tom Thumb. These together with the dominant factor *D* from Tom Thumb resulted in a height growth not greatly different from that of F_1 plants of crosses between Tom Thumb and tall normals. If this supposition is correct, the F_2 plants should exhibit a wide range of variation in height, some of the F_2 normals being taller than the normal Tom Thumb parent and some of the dwarfs being shorter than the dwarf parent. This expectation was strikingly realized. The F_2 normals ranged in height from that of Tom Thumb as ordinarily grown to nearly twice that height with a mean height considerably above that of Tom Thumb. The F_2 dwarfs did not show a great range in height, due perhaps to the small numbers involved, but their mean height was considerably below that of the usual height of dwarfs. While the ranges of heights of the F_2 normals and dwarfs did not overlap, they came close together. Moreover, the shortest F_2 normals were no taller than the tallest dwarfs of other cultures. Evidently the height of these F_2's was influenced by growth factors besides *D d*.

Anther-ear plants also are ordinarily shorter than normal plants of the same cultures. In some cultures anther ear is sharply differentiated from normal by its semi-dwarfness (table 4, culture 294), there being no over-

TABLE 4

Heights of normal and anther-ear plants of F_2 progenies of the cross anther ear × normal.

PEDIGREE CULTURE	HEIGHT OF PLANTS IN DECIMETERS																	
	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	Mean
294 Normal.....								2	5	5	11	4	3	1	1			18.9
Anther ear....	1	5	1	1	1													10.6
298 Normal.....				1		1	3	4	6	8	8	8	8	3	1			18.6
Anther ear....	1	1			2	5	2	1	2	1								14.0
7421 Normal.....						1			4	1	3	13	9	9	5	4	2	21.1
Anther ear....					1	1		3	4	4	1	5	1	1				17.9

lapping of the height distributions of the two forms. In such cultures the principal factor pair for plant height differentiating normal from anther ear is $A_n\ a_n$. Since $a_n\ a_n$ does not exert so strong a dwarfing effect as does *d d*, there is less difference in height between normals and anther ears even of such a culture as 294 than there usually is between normals and dwarfs. In other cultures the anther-ear plants, while on the average shorter than normals, overlap them in height distributions so much that separation of the two forms must be based on characters other than height (table 4, cultures 298 and 7421).

The cause or causes of overlapping in the height ranges of normal and anther-ear plants of some cultures and the sharp segregation in height of these two types in other cultures cannot be determined with certainty from the data now available. Certain possibilities, however, should be noted.

There may exist a series of allelomorphs instead of the single pair A_n a_n. The fact that all our anther-ear stocks came originally from a single plant makes this supposition improbable. Since, however, this plant was not self-pollinated, the suggestion is not an impossible one. Moreover, there is always the possibility of one or more mutations having occurred in one or both members of the A_n a_n pair since the original stock was obtained. If there exist an allelomorph of A_n a_n which has a less dwarfing effect than a_n, just as a_n has less effect than d, the anther-ear plants of some cultures might well be considerably taller than those of other cultures without any corresponding difference in height between the normal plants of the two lots.

A second possibility is that other growth factors may be linked with the A_n a_n pair. If multiple factors affecting height growth are distributed widely throughout the chromosome complex of maize, as seems probable, some of them must necessarily be linked with A_n a_n. Such linkage might well bring about overlapping of heights of normal with those of anther-ear plants in certain cultures and sharp segregation of heights in others, the result in any one culture depending on whether certain dominant or partially dominant growth factors entered the cross with A_n or with a_n. The effects of such linkage might of course be more or less masked by the presence in a heterozygous condition of other growth factors not linked with A_n a_n.

While it seems highly probable that some growth factors may be linked with A_n a_n, resort to this hypothesis is not essential to a plausible explanation of the data presented in table 4. It may well be that in such sharply segregating cultures as 294 few growth factors other than A_n a_n were heterozygous while in such overlapping cultures as 298 and 7421 numerous other growth factors were heterozygous. In case of the latter cultures, such normal plants as happened to receive few dominant or partially dominant growth factors would be expected to be relatively short in stature and such anther-ear plants as happened to receive many of these factors would be relatively tall, thus insuring an overlapping of the height distributions of the two types. If this hypothesis be true, both normal and anther-ear plants should exhibit greater variability in height in overlapping cultures than in sharply segregating ones. While the numbers of

individuals involved were perhaps too few to afford conclusive evidence on this point, it is apparent that the overlapping cultures of table 4 were more variable than the sharply segregating culture.

As bearing further upon this suggestion of the possible effect of heterozygous growth factors in addition to $A_n\ a_n$, a cross of the very small normal race Tom Thumb with a stock of anther ear of relatively low stature may be cited. Unfortunately individual records of height of plants were not made for this cross, our records indicating merely the range of variation in height of the F_2 normals and anther ears. The F_1's of this cross, culture 1848, were normals of medium height such as result from crosses of Tom Thumb with a tall normal race. The ranges in height of the F_2 normals and anther ears are shown in table 5. For comparison the heights of anther-ear plants of culture 294 (table 4), of Tom Thumb plants of culture 508 (table 2), and of dwarf plants of culture 7424 (table 1) are included. The added cultures were not grown at the same time as the F_2 cultures and are not therefore wholly comparable with them.

TABLE 5

Range of heights of parent types and of F_2 normal and anther-ear plants of the cross anther ear × Tom Thumb, and of a culture of dwarf plants.

PEDIGREE CULTURE	HEIGHT OF PLANTS IN DECIMETERS												
	3	4	5	6	7	8	9	10	11	12	13	14	15	16
294 Anther ear (table 4)							←	—	—	—	→			
508 Tom Thumb (table 2)					←	—	—	→						
7424 Dwarf (table 1)	←	—	—	→										
8286 F_2 normal				←	—	—	—	—	—	—	—	—	—	→
F_2 anther ear			←	—	—	—	—	—	—	—	—	—	→	

It is evident from the height distributions shown in table 5 that an unusually large complex of growth factors besides $A_n\ a_n$ were brought into this cross by Tom Thumb pop and the anther-ear parent. The anther-ear plants of F_2 ranged in height from above the usual height of the stock of anther ear used as one parent to much below the ordinary height of Tom Thumb and even to about the average height of ordinary dwarfs. The F_2 normals showed about the same range of heights but were on the whole slightly taller than the anther ears.

From the above it seems safe to conclude that the diverse heights of different anther-ear stocks, and of dwarfs as well, are in large measure due to the influence of growth factors other than $A_n\ a_n$ and $D\ d$. The overlapping in height of anther-ear and normal plants of some F_2 cultures

TABLE 6

Heights of F_2 normal, anther ear, dwarf, and dwarf anther ear of the cross dwarf × anther ear and of a cross equivalent to normal × dwarf anther ear.

PEDIGREE CULTURE	HEIGHT OF PLANTS IN DECIMETERS																									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	Mean
7420 Normal												4	5	3	10	9	4	5	6							15.7
Anther ear									3	2	2	1	1	1		1										11.3
Dwarf			3	9	2	1																				4.1
Dwarf anther ear	1																									
10536 Normal												1		2	3	7	14	7	9	2	8	2	5	2	1	18.6
Anther ear									2		2		2	2	2	6	1		2	1						14.8
Dwarf			4	8	2	1																				
10735, 10736 Normal														1	2	3	5	3	5	6	3	3				18.4
Anther ear											2		1	1			1	2	2							15.6
Dwarf			3	3																						3.5
Dwarf anther ear	1																									

and the sharp height segregation of these types in other cultures appears to be due largely to the relative number of growth factors that happen to be in a heterozygous condition in the several cases. The linkage of certain growth factors with A_n a_n probably has an effect also. Whether allelomorphs of A_n a_n are involved cannot be determined from the data at hand, but that the pairs A_n a_n and D d, particularly the latter, play an important part in determining height of plant is evident. In no single F_2 culture as yet observed have dwarfs overlapped in height either normals or anther ears.

That dwarfs are distinctly smaller than the other two types, normal and anther ear, is particularly well shown in table 6 in which are presented data from F_2 cultures involving all these types together with a few plants of the double recessive dwarf anther ear. One of these cultures was an F_2 of a cross of normal with anther ear heterozygous for dwarf, equivalent to a cross of normal with dwarf anther ear, and the others were F_2's of the cross dwarf $\times$ anther ear (table 13). The normals in all cases were somewhat taller than the anther ears. This increase in height over the parent types, dwarf and anther ear, is doubtless due in the main to the interaction of the dominant factors D and A_n. The sharp segregation in height of dwarfs is in strong contrast with the overlapping of the heights of normal and anther ear. So few dwarf anther ears were measured that it cannot be determined from the data whether dwarf and dwarf anther ear ever overlap in height ranges as do normal and anther ear.

RUDIMENTARY STAMENS IN DWARF AND ANTHER-EAR TYPES

The original stocks of dwarf and of anther ear had well developed stamens throughout the ears. In 1915 in a culture consisting of normals and dwarfs, the ears of some of the dwarf plants were found to have only rudimentary stamens. Since other dwarfs of the same culture had well developed stamens, the character was evidently heterozygous in the stock. A dwarf with rudimentary ear stamens was crossed with an anther ear with normal ear stamens. The F_1 plants had no stamens in their ears or only a few very rudimentary ones such as are sometimes seen in normal stocks. The F_2 cultures of this cross (table 13) consisted of normal, dwarf, anther ear and dwarf anther ear. The latter had no ears and the normals were without ear stamens. Of the F_2 dwarfs and anther ears some had well formed ear stamens and others only rudimentary ones. The 64 dwarf and anther-ear plants were distributed among the four classes as follows:

	Stamens normal	*Stamens rudimentary*	*Total*
Dwarf	11	22	33
Anther ear	21	10	31
Total	32	32	64

While the number of individuals observed is too small for reliable indications, the frequency distribution noted above suggests that a factor for rudimentary stamens in this material may be linked with *d*. But no simple linkage relation of that kind would be expected to give the observed frequency distribution in an F_2 culture. If a factor for rudimentary stamens should be found to behave as a dominant in dwarf and as a recessive in anther ear, the observed F_2 frequencies should be fairly closely realized. Until the question is investigated more fully, no purpose is served by further speculation.

SUMMARY

Two andromonoecious types of maize, dwarf and anther ear, are described and compared with normal maize and with the double recessive, dwarf anther ear. Both types are characterized by the presence of stamens,—usually well developed but in some stocks rudimentary,—throughout the ear. These types have relatively short internodes and relatively short, broad leaves. The tassels are stout and relatively little branched. The ears of both types are short and comparatively thick and frequently end in unbranched tassel-like spikes. With respect to all these characters except perfect flowers the dwarf type is more extreme than the anther-ear type, the latter, in some stocks, approaching closely to the normal type of maize. Dwarf anther ear is an extremely dwarf type with very broad, short leaves and thick, unbranched tassels. No pollen is shed and no ear shoots have been produced by this type.

Both dwarf and anther ear are simple recessives to normal. Intercrosses of these two types give normal plants in F_1 and normals, dwarfs, anther ears, and dwarf anther ears in F_2 in approximately a 9 : 3 : 3 : 1 relation. Owing to failure of the more dwarf forms to germinate or to persist under field conditions, this numerical relation of the several types is approached closely only in greenhouse cultures. This F_2 behavior is interpreted on the basis of two independent factor pairs, $D\ d$ and $A_n\ a_n$. The hypothesis has been tested by the behavior in F_3 and later generations of the several types and by intercrosses between them.

Marked diversities have been observed in height of plants between different anther-ear stocks and to a less degree between different dwarf stocks. These diversities are ascribed to the influence of partially domi-

nant growth factors other than $D\ d$ and $A_n\ a_n$, such as are assumed to differentiate tall and short races of normal maize.

Dwarf plants with rudimentary ear stamens crossed with anther-ear plants with fully developed stamens have given in F_2 both dwarf and anther ear with both rudimentary and well developed ear stamens. The numerical relation of these four types is not that expected on the basis either of independent inheritance or of any simple linkage relation.

LITERATURE CITED

BLARINGHEM, L. 1908 Mutation et traumatismes. 248 pp. Paris: F. Alcan.

EAST, E. M., and HAYES, H. K. 1911 Inheritance in maize. Connecticut Agric. Exp. Sta., Bull. 167, 142 pp.

EMERSON, R. A. 1911 Genetic correlation and spurious allelomorphism in maize. Ann. Rpt. Nebraska Agric. Exp. Sta., **24**: 59–90.

1912 The inheritance of certain abnormalities in maize. Proc. Amer. Breeders' Assoc. **8**: 385–399.

EMERSON, R. A., and EAST, E. M. 1913 The inheritance of quantitative characters in maize. Nebraska Agric. Exp. Sta., Research Bull. 2, 120 pp.

KEMPTON, JAMES H. 1913 Floral abnormalities in maize. U. S. Dept. Agric., Bur. Plant Ind. Bull. 278, 16 pp.

MONTGOMERY, E. G. 1906 What is an ear of corn? Pop. Sci. Monthly **69**: 55–62.

1911 Perfect flowers of maize. Pop. Sci. Monthly **79**: 346–349.

WEATHERWAX, PAUL 1916 Morphology of the flowers of *Zea mays*. Bull. Torrey Bot. Club **43**: 127–144.

1917 The development of spikelets of *Zea mays*. Bull. Torrey Bot. Club **44**: 483–496.

APPENDIX—TABLES

TABLE 7

F_2 progenies of the cross dwarf × normal.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_2 PLANTS			
	F_1	F_2	Field		Greenhouse	
			Normal	Dwarf	Normal	Dwarf
	164 — 4	396	22	2		
	1794— 3	2237	63	19		
	1799— 6	2226	15	3		
	9	2221, 2222	22	11		
	14	2225	16	3		
	28	2224	13	3		
	32	2220	15	5		
	37	2223	12	4		
	1914— 1	1946	37	6		
	2	1952	34	2		
	3	1948	37	3		
	4	1953	49	4		
	5	1947, 1949	70	7		
	8	1945	38	7		
	9	1944	41	3		
	10	1950	25	4		
1	11	1954	47	10		
	12	1951	36	1		
	2403— 1	2962—2966	34	7		
	2967— 2	6826, 6827	71	20		
	10	6830, 6831	38	9		
	11	6828, 6829	112	29		
	4257— 1	5539, 5540	41	16		
	2	5541, 5552	61	14		
	11	5536—5538	7	3		
	9613— 8	10728, 10729	67	15		
	9643— 6	10400	19	3		
	8	10401, 10731	23	7		
	A1400—4	10730	19	5		
	S7—4	7697	11	6		
	Total, 30 progenies		1095	231		

TABLE 7 (*continued*)

F_2 progenies of the cross dwarf × normal.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_2 PLANTS			
			Field		Greenhouse	
	F_1	F_2	Normal	Dwarf	Normal	Dwarf
2	1120— 1	1451			32	15
	2	1452			35	9
	3	1453			39	10
	4	1454			37	12
	3608—18	6780			18	4
	7351— 5	S220—S222			84	23
	7859— 1	S294, S295			31	13
	7867— 1	S289—S291			12	8
	2	S292, S293			15	5
	7869— 2	S298, S299			16	6
	7870— 1	S300			38	9
	7871— 1	S301, S302			30	9
	S6—2	7759, S200, S201			26	8
	6	7760, S202, S203			15	7
	Total, 14 progenies				436	138
3	3609—18	6779, 7408	15	1	30	13
	4257— 3	5535, 6784	62	10	73	27
	21	5543, 6740, 6783	48	19	75	18
	7417— 6	S217	13	5	67	29
	7	S218	6	0	27	8
	7440— 1	S190	30	3	26	12
	2	S191	25	11	37	12
	3	S192	35	1	37	13
	5	S195	32	7	35	13
	S7—7	7434, S189	11	3	103	40
	Total, 10 progenies		277	60	510	185
	Total, 54 progenies		1372	291	946	323

TABLE 8

F_2 progenies of the cross dwarf × normal, back-crossed with dwarf.

PEDIGREE NUMBERS		NUMBER OF F_2 PLANTS	
F_1 × dwarf	F_2	Normal	Dwarf
6785—1 × 6784—18	7429, 7430, S179	37	44
1 × 28	7431, 7432	63	60
S180—1 × S300—3	8857	14	16
S184—4 × S202—1	8858, 8859	89	98
8 × 1	8860—8863	51	46
S186—2 × S294—2	8855, 8856	17	22
S207—6 × A744—11	9643	8	4
S219—2 × S222—7	8865—8868	52	64
Total, 8 progenies		331	354

TABLE 9

F_3 and equivalent F_4 and F_5 progenies of the cross dwarf × normal.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_2, F_4 AND F_5 PLANTS	
			Normal	Dwarf
	F_2	F_3		
	1945–18	2423	13	
	2220–14	4208	30	
	6829– 9	7655–7657	25	
	F_3	F_4		
1				
	5049–13	6913, 6914	11	
	37	6919–6921	78	
	5054–10	6745, 6872	41	
	5055– 2	6871, 7315	10	
	Total, 7 progenies		208	
	F_2	F_3		
	1945–10	2308	8	3
	11	2305, 2306	16	7
	1946–34	2523, 2524	39	11
	1949–43	2535, 2536	12	13
	2966– 7	5049–5055	86	15
	6828–12	7658–7660	24	1
	F_3	F_4		
	5049–25	6816–6818	65	19
2	5050– 1	6874	13	4
	6	6822–6824, 7058, 7059	79	12
	5052– 3	6833	18	6
	7	6825, 7202, 7323, 7442	43	4
	12	6832	21	11
	5053– 1	6875, 6911	37	9
	7660– 1	8536	8	3
	F_4	F_5		
	6832–31	7313, S181	15	2
	Total, 15 progenies		484	120
	F_2	F		
	7429–18	S213, S214		4
	7760– 9	S206		5
	9643– 3	10402, 10732		12
3	S189–50 × S217–27	9644		20
	F_3	F_4		
	5052– 5	S5	1	15
	Total, 5 progenies		1	56

TABLE 10

F_2 progenies of the cross anther ear × normal.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_2 PLANTS			
			Field		Greenhouse	
	F_1	F_2	Normal	Anther ear	Normal	Anther ear
1	19C −2	294−296	38	8		
	19E −2	297−299	59	15		
	419 −3	1019−1021	26	8		
	432 −1	1037	31	6		
	4	1040	146	30		
	2451−1	4089	8	2		
	3	4088	8	3		
	4078−9	6781, 6782	30	15		
	6786−1	7434, 7435				
		S279, S280	63	12		
	5	7436, 7437	12	4		
	7874−1	S282, S283	19	3		
	7875−1	S286	15	2		
	Total, 12 progenies		455	108		
2	792 −4	1439			12	5
	796 −3	1442			23	19
	799 −3	1443			29	9
	800 −2	1445			6	1
	801 −1	1446			32	13
	803 −1	1448			53	16
	2	1447			32	12
	824 −1	1438			36	14
	3	1462			23	7
	S272−2	9763			24	8
	S273−3	9764			29	17
	S275−3	9765			11	3
	S278−1	9766			10	4
	Total, 13 progenies				320	128
3	792 −2	1440, 1847	30	5	25	8
	800 −1	1444, 1764, 1843	23	3	30	16
	Total, 2 progenies		53	8	55	24
	Total, 27 progenies		508	116	375	152

TABLE 11

F_2 and equivalent F_4 progenies of the cross anther ear × normal.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_2 PLANTS	
			Normal	Anther ear
	F_2	F_3		
	295 −1	810−812, 1766	53	
	297 −3	816−820	105	
1	1843−3	2241	114	
	7410−1 × 7436−7	S269	12	
	Total, 4 progenies		284	
	294 − 1	805, 806	33	3
	296 − 2	814	17	5
	297 − 1	821−823, 1839	12	3
	1843− 5	2243	110	35
	11	2242	61	17
2	12	2240	66	16
	1847−16	2244	69	21
	F_3	F_4		
	805 −3	1449	25	10
	Total, 8 progenies		393	110
	294 −28	746, 784, 1123		7
	6781−22	7409		15
3	6782−16	7410	2	14
	Total, 3 progenies		2	36

TABLE 13

F_2 progenies of the cross dwarf × anther ear and crosses equivalent to normal × dwarf anther ear.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_2 PLANTS							
			Field				Greenhouse			
	F_1	F_2	Normal	Dwarf	Anther ear	Dwarf anther ear	Normal	Dwarf	Anther ear	Dwarf anther ear
1	7407–1	S176					25	9	8	0
	5	10735, S177	20	5	8	1	18	11	9	6
	9	10736, S178	25	2	7	0	31	10	8	1
	7418–3	S174	31	4	6	0				
	6787–1	7420, S97	23	7	6	0	39	14	15	5
	S3–3	9754					104	40	34	14
	Total, 6 progenies		99	18	27	1	217	84	74	26
2	9161–5	10536	83	25	22	0				
	S224–2	9761					47	12	13	4
	Total, 8 progenies		182	43	49	1	264	96	87	30

TABLE 12

F_1 progenies of the cross dwarf × anther ear.

PEDIGREE NUMBERS		NUMBER OF F_1 PLANTS NORMAL
P_1	F_1	
2451– 2 × 2440– 2	4084	2
5053– 3 × L188– 1	6787, 6788, S3, S4	29
6779–23 × 6781–22	7407	10
6781– 3 × 6831–14	7418, S175	35
6783–16 × 6781–18	7419	4
7409–14 × 7439–12	S276	3
Total, 6 progenies		83

TABLE 14

F_2 progenies of the cross dwarf × anther ear back-crossed with anther ear and with dwarf.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_2 PLANTS		
	F_1 × Anther ear	F_2	Normal	Anther ear	Dwarf
1	7418– 5 × 7409–7	S167, S168	54	31	
	13 × 7	S169	36	17	
	S1–5 × 6781–27	7421–7423, S98	67	45	
	Total, 3 progenies		157	93	
	F_1 × dwarf				
2	7418–11 × 7439–21	S91, S170, S171	45		16
	13 × 21	S92, S172, S173	128		85
	S4–1 × 6827–10	7427, 7428, S100	25		31
	S5–15 × S4–5	7424–7426, S99	26		30
	S175–15 × S200–2	8864	50		42
	Total, 5 progenies		274		204

TABLE 15

F_3 progenies of the cross dwarf × anther ear.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_3 PLANTS			
	F_2	F_3	Normal	Dwarf	Anther ear	Dwarf anther ear
1	7420–19	9624, S60, S136	155			
	54	9635, S81, S135	135			
	Total, 2 progenies		290			
2	7420– 5	9619, S51, S128	107	28		
	29	9628, S69, S145	53	18		
	33	9631, S74, S150	100	15		
	50	9633, S156	47	10		
	74	9640, S86, S163	67	15		
	Total, 5 progenies		374	86		
3	7420–15	9623, S57, S133	77		21	
	20	9625, S63, S139	90		28	
	35	9632, S77, S153	82		12	
	Total, 3 progenies		249		61	
4	7420– 2	9618, S48, S126	74	19	21	1
	6	9620, S52, S129	63	14	27	
	12	9622, S54, S130	63	19	18	
	21	9626, S64, S140	68	19	18	
	23	9627, S66, S142	20	6	10	
	30	9629, S70, S146	44	15	17	1
	32	9630, S71, S147	58	16	13	
	51	9634, S80, S157	67	16	16	2
	58	9636, S82, S159	75	10	11	
	61	9637, S83, S160	54	8	9	2
	62	9638, S84, S161	45	6	17	
	68	9639, S85, S162	44	7	11	2
	76	9641, S87, S164	41	12	14	
	77	9642, S88, S165	58	7	13	3
	Total, 14 progenies		774	174	215	11
5	7420–34	9757		51		
	S97–14	9775		17		6

TABLE 16

F_3 progenies of the cross dwarf × anther ear back-crossed with anther ear.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_3 PLANTS			
	F_2	F_3	Normal	Dwarf	Anther ear	Dwarf anther ear
1	7421−11	S242	11		2	
	20	S246	12		2	
	7422− 1	S247	24		3	
	3	S248	27		3	
	11	S251	22		5	
	12	S252	22		8	
	28	S255	18		7	
	32	S256	14		2	
	7423− 1	S257	23		5	
	33	S264	6		2	
	S98 − 4	9755	73		24	
	Total, 11 progenies		252		63	
2	7421−19	S245	19	1	5	
	7422− 5	S249	12	1	2	
	7	S250	17	1	2	
	26	S254	25	2	6	
	7423− 4	S258	24	4	4	
	5	S259	14	2	4	
	9	S260	25	6	3	
	18	S262	19	4	6	
	20	S263	21	2	8	
	S99 −10	9756	10	3	3	2
	Total, 10 progenies		186	26	43	2
3	7421−17	S104			9	
	21	S35, S105			13	
	7422− 6	9769, S36, S106			18	
	7423−32	9770, S47			17	
	Total, 4 progenies				57	
4	7422− 8	S37, S107			14	7
	10	S109			10	2
	16	S40, S95			52	17
	21	9768, S41, S112			37	6
	7423− 6	9771, S44			14	5
	Total, 5 progenies				127	37

TABLE 17

F_3 progenies of the cross dwarf × anther ear back-crossed with dwarf.

GROUP	PEDIGREE NUMBERS		NUMBER OF F_3 PLANTS			
	F_2	F_3	Normal	Dwarf	Anther ear	Dwarf anther ear
1	7424—13	S226	26	3		
	19	S228	24	5		
	7426— 1	S231	29	3		
	6	S233	10	3		
	7427— 2	S235	20	7		
	7428— 4	S238	30	3		
	8	S240	22	2		
	Total, 7 progenies		161	26		
2	7425— 4	S230	19	2	2	
	7426— 3	S232	16	2	3	
	8	S234	10	0	2	
	7427— 7	S236	23	1	3	
	7428— 6	S239	25	2	4	
	9	S241	22	3	5	
	Total, 6 progenies		115	10	19	0
3	7427—12	9772, S208		29		
	21	9773, S209, S210		56		
	7428—19	9774, S212		88		
	Total, 3 progenies			173		

PRODUCTION OF HIGH-PROTEIN MAIZE BY MENDELIAN METHODS[1]

H. K. HAYES

Minnesota Experiment Station, St. Paul, Minnesota

Received July 15, 1921

TABLE OF CONTENTS

INTRODUCTION 237
Review of previous studies 238
Materials and methods 239
Protein content in relation to cultural conditions 339
Relation between number of seeds per ear and percentage protein content 242
Comparison of protein content of self-pollinated and wind-pollinated maize 248
Inheritance of protein content 251
Isolation of high-protein races through self-fertilization 252
The use of self-fertilized lines in maize improvement 252
LITERATURE CITED 257

INTRODUCTION

Some plant breeders have preferred to attack crop-improvement problems by empirical methods while others have thought it first necessary to learn the genetics of important crop characters. Thus we have such studies as the Illinois continuous-selection experiments for the purpose of producing maize races which differ in the greatest possible degree in certain chemical constituents such as high and low oil and high and low protein. Experiments of this nature show very clearly the possibilities of plant breeding. They have been widely quoted as illustrations of the creative power of selection. While they show the possibilities of selection they do not materially increase our knowledge of the genetics of maize.

An appreciation of the necessity of a knowledge of the genetics of maize led Dr. E. M. EAST to start an investigation on the physiology of maize inheritance at the CONNECTICUT EXPERIMENT STATION. The present writer had a part in these investigations during the five-year period from

[1] One of several papers submitted to the Faculty of the Bussey Institution of HARVARD UNIVERSITY in partial fulfillment of the requirements for the degree of Doctor of Science. Published with the approval of the Director as paper No. 266 of the Journal Series of the MINNESOTA AGRICULTURAL EXPERIMENT STATION. The helpful suggestions of Dr. E. M. EAST have been of material value in the studies of inheritance of protein which are reported in this paper.

1910 to 1914, inclusive. During this period the work was carried on in coöperation with Dr. EAST. From 1915 until the present time studies in maize breeding have been carried on by the writer at the MINNESOTA EXPERIMENT STATION.

The view of learning the genetics of important economic characters and the building of improved maize varieties through the application of these principles has been the main aim at Minnesota. One phase, therefore, of the Minnesota work has been the isolation of self-fertilized lines and their recrossing.

Protein content is of much interest in the light of the Illinois studies and the writer's experience at Connecticut showed that the genetics of protein content was a complex subject. It was therefore decided to continue this investigation with Minnesota varieties. This plan was followed, not because of the belief that high-protein races were necessarily of great economic importance, but because it was thought that a knowledge of the correct genetic methods of obtaining good-yielding, high-protein races would have a direct bearing on the breeding of maize varieties which excel in other important economic characters.

REVIEW OF PREVIOUS STUDIES

A preliminary report which indicated that high-protein maize could be produced by the isolation through self-fertilization of high-protein races and their subsequent crossing was made by HAYES and GARBER (1919). The main purpose of this paper was to show that the isolation of self-fertilized lines and their subsequent crossing was a logical method for the production of improved economic varieties. Since the publication of this preliminary note, JONES (1920) has discussed the self-fertilization method of corn improvement in some detail. The views of the writer do not materially differ from those of JONES with one exception. JONES states that the recrossing of selfed lines and subsequent field selection will probably lead to the production of a variety which does not materially differ from the normal variety from which the self-fertilized lines were obtained. As all corn varieties contain many abnormalities, the elimination of these and the subsequent crossing of lines which seem of most promise seems a promising method of producing an improved variety.

EAST and JONES (1921) have recently presented an interesting paper on the genetics of protein content and have reviewed much of the literature in this field. The reader is referred to this paper and to those of JONES (1920) and HAYES and GARBER (1919) for reviews of literature in relation to the problem.

The paper by EAST and JONES (1921) will be discussed later in relation to the facts presented in this paper. The present study helps in a measure to clear up some further points in relation to EAST and JONES's discussion. Much of the data collected at Minnesota will be presented only briefly in summarized form for the purpose of showing results with a different series of varieties.

MATERIALS AND METHODS

Standard Minnesota varieties have been used in this problem. Methods of controlling pedigrees and the production of crossed and selfed seed have been carried on in much the same way as at Connecticut. For this reason it does not seem necessary to further discuss these phases.

One of the difficulties of maize breeding by artificial-pollination methods is the possibility of heavy showers which wet the tassel bags and make pollination difficult. A plan used with advantage under Minnesota conditions is to immediately replace all partially damaged bags after each rain.

Two men work together at pollination, one caring for the bag which covers the silk and the other removing the tassel bag and carefully dusting the pollen over the silk. Various men have assisted in this work: P. J. OLSON and M. N. POPE during 1915 and 1916, R. J. GARBER 1917 to 1920, and FRED GRIFFEE in 1920. The analyses have been made by the Division of Agricultural Biochemistry, the percentage figures for protein content being on a dry basis. All analyses were made in duplicate and officially recognized analytical methods were followed.

Although recognizing that the individual seed is the genetic unit and not the plant, nevertheless it was apparent that the protein content of the seeds of an ear are not greatly modified by the immediate effect of cross-pollination. The analysis of an ear represents therefore the development of protein which is possible in the light of the genetic factors carried by the mother plant and under the environmental conditions under which the plant develops.

The effect of environment and the conditions under which an artificially pollinated ear develops in relation to protein content will be first discussed.

PROTEIN CONTENT IN RELATION TO CULTURAL CONDITIONS

EAST and JONES (1921) present analyses compiled by JENKINS and WINTON showing the percentage protein content of the five subspecific maize groups. These results show that the groups do not differ very widely

in average protein content. On the other hand, the range of variation for different analyses showed a content as low as 7.7 percent in one analysis of flint corn while one analysis of sweet corn showed a content as high as 17 percent. EAST and JONES compared analyses of self-fertilized upper and lower ears for six different plants. In general there was a close agreement between the protein content of the upper and lower ear of the same plant. This shows that position on the plant does not widely modify the percentage protein content.

In 1919 three normal corn varieties and one twelve-year self-fertilized strain of Red Flint were grown under different cultural conditions. In the Plant Pathology plot the environmental conditions were favorable and a heavy application of manure was made. The Plant Breeding plot appeared to lack a normal supply of food nutrients and likewise suffered from lack of moisture. Therefore the plants did not develop very vigorously and many of the leaves showed a light green color about tasseling time. Self-fertilized ears under the favorable and unfavorable conditions were analysed for protein content (see table 1).

TABLE 1

Percentage protein content under favorable and unfavorable cultural conditions.

VARIETY	PLACE GROWN	NUMBER EARS	RANGE	AVERAGE NUMBER SEEDS PER EAR	AVERAGE PROTEIN PERCENTAGE	PROTEIN P. PATHOLOGY MINUS PROTEIN P. BREEDING
King Phillip	Plant Pathology	10	12.03–14.78	228	13.64	
	Plant Breeding	15	7.83–14.09	224	11.28	+2.36
Rustler	Plant Pathology	9	9.16–14.82	296	12.35	
	Plant Breeding	4	8.97–12.81	308	10.94	+1.41
Red Flint selfed 12 years	Plant Pathology	39	12.73–14.70	112	13.67	
	Plant Breeding	7	12.69–13.86	91	13.36	+0.31
Minnesota No. 13	Plant Pathology	10	10.91–14.38	471	13.08	
	Plant Breeding	5	8.75–13.75	400	11.51	+1.47

The average number of seeds per ear was about the same for the ears produced in the Plant Pathology and Plant Breeding plots. All three of the normal varieties showed much higher average percentage protein content under the favorable conditions of the Plant Pathology plot. The twelve-year self-fertilized strain of Red Flint gave about the same percentage content of protein under the two conditions.

Results in 1920 again show that protein content is very markedly influenced by cultural conditions. Ten ear samples obtained from normally pollinated Minnesota No. 13 were taken from each of two rows of Minnesota No. 13. Border rows of Minnesota No. 13 were grown which surrounded each row from which ten ear samples were taken. The ten ears for each sample were from consecutive plants in the row, the plants being spaced eighteen inches apart in the row. Results were as follows:

Row	*Protein content percentages ten-ear samples*						
96	10.02	10.19	10.54	11.25	10.62	10.65	10.78
108	12.15	11.31	11.94	11.19	11.15	11.87	11.75

It is apparent that the samples from row 108 are uniformly higher in protein content than the samples from row 96. Thus, as with most economic characters, it is impossible to determine the genetic characters of a plant without growing and examining its progeny.

TABLE 2

Correlation between percentage protein content of self-fertilized ears grown in 1919 and average percent protein in the progeny, 1920.

1919 protein content selfed ears	Average percentage of protein in the progeny 1920												
	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	
8.5							1		1				2
9.0					1		1	1					3
9.5							1						1
10.0								1					1
10.5							2	2					4
11.0						1	1		1		1		4
11.5								2					2
12.0				1	1	2	1			2			7
12.5	1				2			1	1				5
13.0					2			2		2			6
13.5				1	2	1	1	3			2	1	11
14.0							2	1		3	2		8
14.5									1				1
15.0				1					1			1	3
15.5										1			1
	1	0	0	3	8	4	10	13	5	8	5	2	59

$r = +.190 \pm .085$

Although it is a comparatively easy matter to isolate high-protein races through self-fertilization, there is not a very close agreement between the percentage protein content of self-fertilized ears of a variety and the protein content of the resultant progenies (see table 2).

The calculated coefficient of correlation for percentage protein content of self-fertilized ears of normal varieties and the percentage content of the progeny was $+.190 \pm .085$. A part of the lack of relation may be due to differences in number of seeds per ear, which markedly affects the protein-content percentage. One reason for the low coefficient is probably that the conditions under which a plant develops so strongly modify the expression of its inheritance.

RELATION BETWEEN NUMBER OF SEEDS PER EAR AND PERCENTAGE PROTEIN CONTENT

A casual observation showed a strong relation between the number of seeds produced by a self-fertilized ear and percentage of protein produced by the seeds. Results for the correlation between number of seeds per ear and protein content for first-year self-fertilized ears of four normal varieties, for self-fertilized ears of two F_1 crosses between isolated self-fertilized high-protein races, for two strains self-fertilized for five years, and for two strains self-fertilized for many years, are given in tables 3 to 12 inclusive.

TABLE 3

Correlation between number of seeds per ear and protein content in first-year selfed ears of King Phillip Flint.

Percentage protein content	Number of seeds per ear: 75	100	125	150	175	200	225	250	275	300	
10.5							1				1
11.0				1				1			2
11.5					1						1
12.0	1					2					3
12.5				3	2	3				1	9
13.0			1								1
13.5		1	1	3			1				6
14.0		1			1	1					3
14.5	1			1							2
15.0							1				1
15.5				1							1
	2	2	2	9	4	6	3	1	0	1	30

$r = -.303 \pm .112$

TABLE 4

Correlation between number of seeds per ear and protein content in first-year selfed ears of Mercer Flint.

Percentage protein content	Number of seeds per ear									
	75	125	175	225	275	325	375	425	475	
11.5					1	1				2
12.0		1		1	1					3
12.5		1		2				1	1	5
13.0	1			3	2					6
13.5		1		2		1				4
14.0		1		1						2
14.5					2					2
15.0			1		1					2
15.5										0
16.0		1								1
	1	5	1	9	7	2	0	1	1	27

$r = -.254 \pm .121$

TABLE 5

Correlation between number of seeds per ear and protein content in first-year self-fertilized ears of Minnesota No. 13.

Percentage protein content	Number of seeds per ear									
	75	125	175	225	275	325	375	425	475	
10.0				1						1
10.5										0
11.0										0
11.5		1								1
12.0			1		1		1			3
12.5			1		1					2
13.0			1	1	2	1				5
13.5			1	2	1	1			1	6
14.0		1	1	1		1				4
14.5		2	1	1						4
15.0	2	1	1							4
15.5		1		1						2
	2	6	7	7	5	3	1	0	1	32

$r = -.357 \pm .104$

TABLE 6

Correlation between number of seeds per ear and protein content in first-year selfed ears of Rustler Dent.

Percentage protein content	Number of seeds per ear: 125	175	225	275	325	375	
10.0						2	2
11.5				1			1
12.0		1					1
12.5			3	4	2		9
13.0		3			1		4
13.5		2	1				3
14.0	1	1		1			3
14.5		1					1
15.0							0
15.5			1				1
	1	8	5	6	3	2	25

$r = -.643 \pm .104$

TABLE 7

Correlation between number of seeds per ear and protein content in Minnesota No. 13 strain, 13-46-227-1-6, which has been self-fertilized for five years.

Percentage protein content	Number of seeds per ear: 75	125	175	225	
14.5				1	1
15.0			1		1
15.5			1	1	2
16.0	1	1			2
16.5		1	1		2
17.0		1	3		4
17.5	1				1
18.0		1			1
	2	4	6	2	14

$r = -.536 \pm .128$

TABLE 8

Correlation between number of seeds per ear and protein content in self-fertilized ears of related strains, 13-30-211-10-6, -10-7, -10-18, and -10-3, which have been self-fertilized for five years.

Percentage protein content	Number of seeds per ear: 125	175	225	275	325	375	425	475	525	
11.0						1				1
12.0		1								1
12.5										0
13.0				2	1	1				4
13.5				1						1
14.0		1		1	1					3
14.5		2		2		3		1		8
15.0		1	2	1	2	2		1	1	10
15.5		2		1	1	2		1		7
16.0		2	1			1	2	1		7
16.5	2	4	1	1		1	3			12
17.0		2	2			1				5
17.5		2	2	2		1				7
18.0										0
18.5		1		1						2
	2	18	8	12	5	13	5	4	1	68

$r = -.198 \pm .079$

TABLE 9

Correlation between number of seeds per ear and protein content in selfed ears of F_1 crosses, 13-30-211-13 × 13-46-226-1 and 13-30-211-7 × 13-46-226-2.

Percentage protein content	Number of seeds per ear: 25	75	125	175	225	275	325	375	425	
10.5						1				1
11.5							1			1
12.0				1						1
12.5							2		1	3
13.0									2	2
13.5						3				3
14.0						1				1
14.5				2	2	2	1			7
15.0	1		1	1	2	2		2		9
15.5	1	1		8	1	3	1			15
16.0	1	1	2	3	2	1				10
16.5		1	1							2
	3	3	4	15	7	13	5	2	3	55

$r = -.553 \pm .063$

TABLE 10

Correlation between number of seeds per ear and protein content in selfed ears of F_1 crosses 13-46-227-1 × 13-30-211-8, 13-46-227-2 × 13-30-211-9 and 13-30-211-9 × 13-46-227-1.

Number of seeds per ear

Percentage protein content	25	75	125	175	225	275	325	375	
12.5							1		1
13.0						1	1		2
13.5						1		1	2
14.0							1		1
14.5						1	1	1	3
15.0						2	2	1	5
15.5				1	1	3	2		7
16.0			2	2	1	3	7	1	16
16.5			3	5	2	4	1		15
17.0	2	1	2	3	1		1		10
17.5			1	1	1				3
18.0			1						1
	2	1	9	13	6	15	17	4	66

$r = -.588 \pm .054$

TABLE 11

Correlation between number of seeds per ear and protein content in self-fertilized ears of yellow dent strain, 1-6-1, etc., selfed twelve years.

Number of seeds per ear

Percentage protein content	75	125	175	225	275	325	375	
14.0	1	1				1		3
14.5	2	1	1			1		5
15.0		2	1	2	1		1	7
15.5		1						1
16.0			1			1		2
	3	5	3	2	1	3	1	18

$r = +.228 \pm .151$

TABLE 12

Correlation between number of seeds per ear and protein content in self-fertilized ears of Red Flint strain selfed twelve years.

Percentage protein content	Number of seeds per ear: 25	75	125	175	225	
12.5		1	1			2
13.0	2		4			6
13.5	1	7	7	2		17
14.0	1	5	7	2	2	17
14.5	2		1	1		4
	6	13	20	5	2	46

$r = +.105 \pm .098$

Summarized coefficients of correlation are given in table 13.

It is apparent from these results that protein-content percentage is very strongly modified by the number of seeds produced per ear. The correlations are uniformly large for the two F_1 crosses and consistent correlations are obtained with the normal varieties. The lack of any marked correlations for number of seeds per ear and protein content for the strains which have been selfed for many years (except for 13-46-227-1-6) are exceptions to the general rule. In this respect the results are in agreement with the effect of cultural conditions on protein content, for the protein percentages

TABLE 13

Coefficients of correlation for percentage protein content in relation to number of seeds per self-fertilized ear.[2]

STRAIN OR VARIETY	HISTORY	NUMBER EARS	CORRELATION COEFFICIENT
King Phillip Flint	Normal variety	30	$-.303 \pm .112$
Mercer Flint	Normal variety	27	$-.254 \pm .121$
Minnesota No. 13 Dent	Normal variety	32	$-.357 \pm .104$
Rustler Dent	Normal variety	25	$-.643 \pm .104$
Minnesota No. 13, selfed	Selfed 5 years	14	$-.536 \pm .128$
Minnesota No. 13, selfed	Selfed 5 years	68	$-.198 \pm .079$
F_1 Cross, No. 13 strains	High × High	55	$-.553 \pm .063$
F_1 Cross, No. 13 strains	High × High	66	$-.588 \pm .054$
Leaming, 1−6−1, etc.	Selfed 12 years	18	$+.288 \pm .151$
Red Flint, 69−6, etc.	Selfed 12 years	46	$+.105 \pm .098$

[2] Mr. A. N. WILCOX assisted in checking these correlation coefficients.

of the ears of the Red Flint strain which had been self-fertilized for twelve years were about the same for plants growing under unfavorable as under favorable conditions.

COMPARISON OF PROTEIN CONTENT OF SELF-POLLINATED AND WIND-POLLINATED MAIZE

Results obtained in the early years of study showed that self-pollinated ears were markedly higher in percentage protein content than normally pollinated ears. This was thought at first to be due largely to some change in light or temperature conditions which resulted from covering the self-fertilized ears with a paper bag. A later explanation advanced was given by EAST and JONES (1921), who believe a part of the difference in percentage protein content between self-pollinated and normally pollinated seeds is due to the immediate heterotic effect of cross-pollination. The best sort of evidence to show the immediate effect of cross-pollination is a comparison of selfed and cross-pollinated seeds of the same ear. Results of this nature are discussed in some detail by EAST and JONES who present a total of 52 analyses in which comparisons are made of self- and cross-pollinated seeds from the same ears. Averaging the results presented by EAST and JONES gives a protein content of 11.61 percent for cross-pollinated seeds and 11.76 percent for self-pollinated seeds. This is a reduction of 1.3 percent in protein content as an immediate effect of cross-pollination. Nine ears from similar pollinations, in which low and medium protein strains were used, were studied at Minnesota. The percentage protein content of the self-fertilized seeds was 15.42 and of the cross-pollinated seeds on the same ears 15.48. These differences are very small in relation to the large differences shown by EAST and JONES between open-pollinated and self-pollinated ears.

A fact of some interest not brought out by EAST and JONES would seem to show that the covering of the ear with a bag does not in itself affect the percentage protein content of the seeds produced by such a covered ear. To illustrate this point open-pollinated (O.P.) and selfed ears (S.) of the 1919 crop season are compared for protein content (see table 14).

Very marked differences in protein content for open-pollinated and self-pollinated seeds are shown for ears of F_1 crosses between high- and low-protein races, a difference of over 50 percent being obtained in one case. Of four normal varieties, three gave markedly lower protein content in open-pollinated than in self-pollinated ears. Without discussing the other results in detail we may briefly note the strains which have been selfed for several years. Of five strains of Minnesota No. 13 self-fertilized for

TABLE 14

Comparison of protein content of open-pollinated and self-fertilized ears in relation to number of years that the strain has been selfed. Grown in 1919.

VARIETY AND STRAIN NUMBER	SELFED OR OPEN-POLLINATED	NUMBER OF YEARS SELFED	NUMBER SELFED EARS	PROTEIN CONTENT	DIFFERENCE S.—O.P.	INCREASE PROTEIN CONTENT S. OVER O. P. PERCENT
High Protein × Low Protein, F_1	S.	1	6	13.62		
	O.P.	0		8.68	4.94	58.3
High Protein × Low Protein, F_1	S.	1	4	10.99		
	O. P.	0		8.49	2.50	29.4
High Protein × Low Protein, F_1	S.	1	3	12.28		
	O. P.	0		8.49	3.79	44.6
Longfellow	S.	1	10	13.67		
	O. P.	0		14.11	−.44	−3.1
King Phillip	S.	1	15	12.38		
	O. P.	0		9.14	3.24	35.4
Minnesota No. 13	S.	1	6	11.51		
	O. P.	0		8.72	2.79	32.0
Rustler 31	S.	1	4	10.74		
	O. P.	0		8.95	1.79	20.0
Mercer 33—18	S.	2	12	12.68		
	O. P.	1		9.45	3.23	34.2
Mercer 33—27	S.	2	7	12.20		
	O. P.	1		10.77	1.43	13.3
King Phillip 30—26	S.	2	5	13.83		
	O. P.	1		12.50	1.33	10.6
King Phillip 30—4	S.	2	7	13.00		
	O. P.	1		11.37	1.63	14.3
Rustler 30—23	S.	2	6	13.38		
	O. P.	1		10.30	3.08	29.9
Minnesota No. 13 13—26	S.	2	7	11.67		
	O. P.	1		9.27	2.40	25.5
Minnesota No. 13 13-3	S.	2	3	13.01		
	O. P.	1		10.55	2.46	23.3

TABLE 14 (continued)

VARIETY AND STRAIN NUMBER	SELFED OR OPEN-POLLINATED	NUMBER OF YEARS SELFED	NUMBER SELFED EARS	PROTEIN CONTENT	DIFFERENCE S. — O.P.	INCREASE PROTEIN CONTENT S. OVER O. P. PERCENT
Minnesota No. 13 13–22	S.	2	10	13.29		
	O. P.	1		11.99	1.30	10.8
Minnesota No. 13 13–23	S.	2	2	12.19		
	O. P.	1		12.18	.01	0.0
Minnesota No. 13 13–24	S.	2	3	13.01		
	O. P.	1		10.13	2.88	28.4
Minnesota No. 13 13–46–227–1–6	S.	5	14	16.39		
	O. P.	4		14.94	1.45	9.7
Minnesota No. 13 13–30–211–10–3	S.	5	14	15.45		
	O. P.	4		15.87	−.43	−2.6
Minnesota No. 13 13–30–211–10–18	S.	5	18	15.52		
	O. P.	4		15.15	.37	0.0
Minnesota No. 13 13–30–211–10–6	S.	5	20	15.45		
	O. P.	4		16.74	−1.29	−7.7
Red Flint	S.	12	7	13.36		
	O. P.	11		13.68	−.32	−2.3

five years, three gave a higher protein content in open-pollinated than in self-pollinated ears. Likewise the twelve-year self-fertilized strain of Red Flint gave a higher protein content in the open-pollinated ears though the differences were very small. These facts seem to show that covering an ear with a bag does not in itself greatly modify protein content.

The facts presented here and the data of EAST and JONES lead to the conclusion that the main cause of the difference in protein content of self-fertilized and normally pollinated ears is the marked correlation between the percentage protein content of an ear and the number of seeds produced by the ear. This fact deserves careful consideration in an analysis of the effects of continuous selection in self-fertilized lines. On the average there is a gradual decrease in yield in self-fertilized lines of maize which for a considerable number of years or until homozygosis is reached is roughly proportional to the number of years which the strain has been self-fertilized. During the period of this reduction in yield there is a corresponding reduction in the average number of seeds produced by either self-pollinated or

normally pollinated ears. This leads naturally, when one considers the correlations previously noted, to a gradual increase in percentage protein content.

INHERITANCE OF PROTEIN CONTENT

It has been noted previously that the percentage protein content of the seeds of an F_1 ear frequently closely corresponds to the percentage protein content of the lower-protein parent. It was also demonstrated by HAYES and GARBER (1919) that F_1 ears of crosses between self-fertilized high-protein races have considerably lower percentage protein content than the parental self-fertilized lines. These facts and a considerable body of similar evidence presented by EAST and JONES led these investigators to conclude that although "there was a semblance of dominance of low protein" over high protein, this was largely due to the effect of heterosis. This apparent dominance has, however, much value from the breeding standpoint. Whether the apparent recessiveness of high protein is due to genetic factors for protein content or partly due to the effect of other growth factors does not detract from the value of the relation between protein content of a self-fertilized ear and the protein content of its progeny. As high protein apparently behaves in inheritance as a recessive character, all that is necessary in order to isolate high-protein races is to practice self-fertilization and then determine which lines breed true for higher protein content by an analysis of the average of the progeny. Even though continued selection isolates higher-protein races, this may not always be due to the isolation of genetic factors for a greater synthesis of protein but due to the elimination of growth factors and the consequent reduction in size of ear and seed which results in a higher percentage protein content.

Crosses between high- and medium-protein races which had been isolated through self-fertilization were made at Minnesota in 1918. A considerable number of F_2 ears were grown in 1920 and normal pollination was permitted. The percentage protein content of the F_2-generation ears was compared with the content of ears of Minnesota No. 13 and with self-fertilized ears of the long-time-selfed strain of Red Flint which was grown in 1919 (see table 15).

The F_2 generation was about as variable for percentage protein content as normal ears of Minnesota No. 13. The low coefficient of variability of the Red Flint ears is of considerable interest. Comparatively homozygous lines for protein content may therefore be isolated, since segregation does occur in those generations which are commonly known as the segre-

gating generations. Protein content is therefore inherited in much the same way as other characters which are dependent for their full expression on many different inherited factors of the plant and likewise upon environmental conditions.

TABLE 15

Comparison of variability for percentage protein content of normally pollinated Minnesota No. 13 ears, F_2 open-pollinated ears of a cross of Minnesota No. 13 high-protein strain × Minnesota No. 13 medium-protein strain, and self-fertilized ears of a twelve-year self-fertilized strain of Red Flint (1920).

SOURCE	PROTEIN PERCENTAGE CLASSES											TOTAL	MEAN	C. V.
	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0			
Normal Minnesota No. 13	7	8	12	19	18	10	12	2					88	12.2±.07	7.46±.38
F_2 cross		2	3	9	12	21	16	11	7	4	2	1	88	13.3±.07	7.59±.39
12-year self-fertilized Red Flint					2	6	17	17	4				46	13.7±.05	3.50±.25

ISOLATION OF HIGH-PROTEIN RACES THROUGH SELF-FERTILIZATION

EAST and JONES have given considerable evidence to show the ease with which high-protein races may be isolated through self-fertilization. For this reason Minnesota data will be only briefly noted. The facts are presented solely for the purpose of showing that high-protein races may be isolated in standard Minnesota varieties (see table 16).

It hardly seems necessary to emphasize these results by extended discussion. The results corroborate those of EAST and JONES and show the value of the method. Perhaps it may serve a useful purpose to point out that similar methods seem the logical means of isolating self-fertilized families which are resistant to plant diseases such as corn smut and root rot. Chlorophyll degeneracies which are prevalent in nearly all commercial varieties may be eliminated by similar means. The idea deserves wide consideration in the hands of plant breeders interested in maize. If all experimental stations in the corn belt could vigorously attack the problem and if coöperation between investigators could be more fully developed the results would be of undoubted value to the corn-growing industry.

THE USE OF SELF-FERTILIZED LINES IN MAIZE IMPROVEMENT

Several methods of the utilization of self-fertilized lines suggest themselves. JONES's (1920) suggestion of the double cross seems worthy of extended trial. The practical difficulties of the method, while not so

great as the use of F_1 crossed seed of two self-fertilized lines (a method originally outlined by SHULL), are due to the necessity of producing four

TABLE 16

Isolation of high-protein strains by self-fertilization.

STRAIN NUMBER	NUMBER YEARS SELFED	YEAR GROWN	PROTEIN CONTENT OF PARENT	PROTEIN CONTENT BULK ANALYSIS OF PROGENY
Minnesota No. 13 13–30	1	1916	16.52	14.47
Minnesota No. 13 13–46	1	1916	16.66	15.82
Minnesota No. 13 13–12	1	1916	16.99	14.10
Minnesota No. 13				
13–30–211	2	1917	16.90	13.06
13–46–226	2	1917	——	12.44
13–46–227	2	1917	17.16	14.03
13–12–181	2	1917	17.40	10.87
Minnesota No. 13				
13–30–211–10	3	1918	——	14.88
13–46–227–1	3	1918	15.10	15.31
13–46–227–3	3	1918	12.72	14.88
13–12–181–2	3	1918	15.28	15.81
Minnesota No. 13				
13–30–211–10–6	4	1919	17.19	15.88
13–30–211–10–14	4	1919	17.19	15.45
13–30–211–10–18	4	1919	17.69	15.15
13–30–211–10–3	4	1919	16.50	15.87
13–46–227–3–6	4	1919	16.19	14.94
13–46–227–1–8	4	1919	15.31	——
13–12–181–2–1	4	1919	15.69	——
Minnesota No. 13				
13–46–227–3–6–26	5	1920	15.59	17.87
13–46–227–3–6–20	5	1920	16.50	17.62
13–30–211–10–6–6	5	1920	16.73	18.12
13–30–211–10–6–10	5	1920	16.57	16.59
13–30–211–10–14–9	5	1920	17.16	17.90
13–30–211–10–14–13	5	1920	18.43	18.22
13–30–211–10–18–22	5	1920	15.20	17.46
13–30–211–10–18–25	5	1920	17.47	16.90
13–12–181–2–1–1	5	1920	14.63	16.44
13–12–181–2–1–9	5	1920	16.72	16.45
13–12–181–2–1–14	5	1920	14.81	18.00
Rustler 31–26–6	2	1920	14.70	17.22
King Phillip 30–4–3	2	1920	13.91	17.50
Mercer 33–27–1	2	1920	15.00	15.03

strains annually and the making of three crosses before obtaining seed for commercial planting.

EAST and JONES point out a logical means of improving strains and show that for high protein the method has proven valuable. This is the isolation of high-protein races, the cross of these races and the subsequent selection in the segregating generations of those self-fertilized ears which

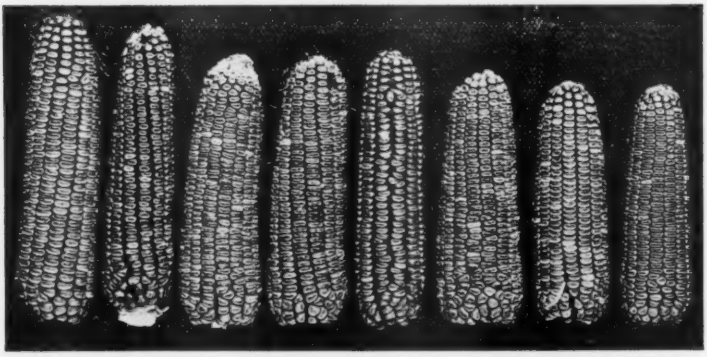

FIGURE 1.—Average ears of Minnesota No. 13 grown in comparison with the double cross of four high-protein strains. 1920. (Photo by HORTON.)

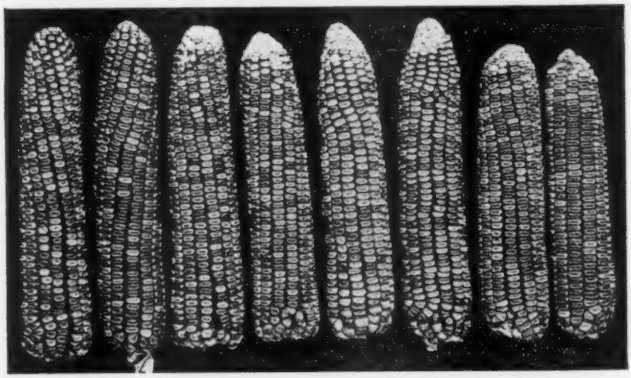

FIGURE 2.—Double cross of four high-protein self-fertilized strains. Same focus as above ears of Minnesota No. 13. 1920. (Photo by HORTON.)

excel for protein content. This is the Mendelian plan which has been adopted by several breeding centers as a means of increasing the yield or improving the economic characters of self-fertilized plants such as our small grain crops.

Another method deserves trial by maize breeders. After several strains have been isolated they may be crossed and field selection practiced for desirable stalks and ears. Such a synthetically produced variety should in some cases, depending on the value of the combination obtained from crossing the selfed strains, give a variety of maize which is a great improvement over the normal variety from which the selfed strains were originally obtained.

FIGURE 3.—Comparative vigor of Minnesota No. 13, left, and the double cross of four self-fertilized high-protein strains at the right. 1920. (Photo by HORTON.)

A double cross of four high-protein strains was tested in 1920. One of the original crosses was made by JONES of Connecticut who used a high-protein strain of Leaming and one of Illinois High Protein. The cross between two Minnesota No. 13 high-protein strains was crossed with the F_1 cross made by JONES and the double cross grown in replicated plot tests in 1920 and compared with Minnesota No. 13. The double cross was slightly later than Minnesota No. 13 but it produced ears very freely and looked very promising. Due to dry weather late in the season the Minnesota No. 13 filled somewhat better than the double cross. Compara-

tive yields and protein content of Minnesota No. 13 and of the double cross are given in table 17 (see figures 1 to 3).

TABLE 17

Comparison of protein content, yields and shelling percent of Minnesota No. 13 and reciprocal double crosses.

SOURCE	SHELLED CORN PER STALK GRAMS	SHELLING PERCENT	PERCENTAGE PROTEIN CONTENT
Double cross	128.2	77.0	14.92
Reciprocal	141.9	76.0	15.14
Minnesota No. 13	144.9	80.6	12.37

An attempt was started in 1916 to produce synthetically a high-protein variety of corn by crossing isolated high-protein races of Minnesota No. 13.

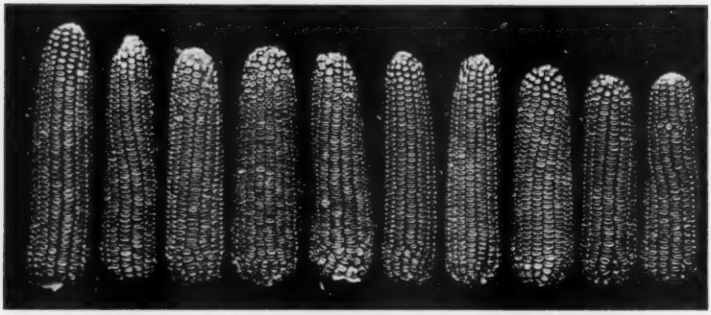

FIGURE 4.—Average ears of Minnesota No. 13. Isolated pasture plot. 1920. (Photo by HORTON.)

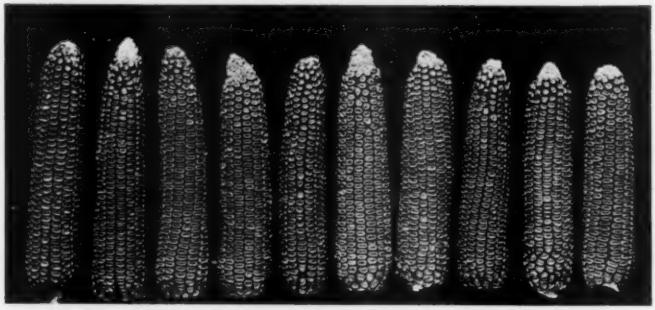

FIGURE 5.—Average ears of High Protein Minnesota No. 13. Produced by crossing two high protein strains and subsequent selection. Isolated pasture plot, 1920. Same reduction as figure 1. (Photo by HORTON.)

The original cross was made between two apparently high-protein strains. In the F_2 generation, however, lethal white chlorophyll seedlings were obtained. F_3 lines were grown from self-fertilized F_2 ears and reciprocal crosses within a line and between different lines were made. Only those families were used which were homozygous for an absence of the white-seedling degeneracy. Seeds of those F_3 reciprocal crossed ears which showed high protein content in both parents, were mixed together and used to plant an isolated plot. A determination of the yield and protein content of this isolated plot was made and compared with a similar plot of Minnesota No. 13 (see table 18 and figures 4 and 5).

TABLE 18

Comparison of yields and protein content of normal Minnesota No. 13 and a synthetically produced Minnesota No. 13 obtained by crossing high-protein strains.

SOURCE	YIELD EAR-CORN PER STALK, GRAMS	PROTEIN CONTENT PERCENTAGE
Normal Minnesota No. 13.....	122.6	12.01
High Protein No. 13.....	122.6	14.99

The results indicate the value of the method. These studies have led to the attempt to breed maize at Minnesota for economic characters other than protein by an adoption of a Mendelian plan which, except for the details of the work, is not essentially different from the method which is being used for small grains and which has already proved its value through actual test.

LITERATURE CITED

EAST, E. M., and JONES, D. F., 1921 Genetic studies on the protein content of maize. Genetics **5**: 543–610.

JONES, D. F., 1920 Selection in self-fertilized lines as a basis for corn improvement. Jour. Amer. Soc. Agron. **12**: 77–100.

HAYES, H. K., and GARBER, R. J., 1919 Synthetic production of high-protein corn in relation to breeding. Jour. Amer. Soc. Agron. **11**: 309–318.

CORRELATION, CAUSATION AND WRIGHT'S THEORY OF "PATH COEFFICIENTS"[1]

HENRY E. NILES

Received August 5, 1921

TABLE OF CONTENTS

INTRODUCTION 258
The method of path coefficients 261
The mathematics of path coefficients 263
WRIGHT'S guinea-pig example 263
Test of WRIGHT'S method 264
Example 1 268
Example 2 279
CONCLUSION 272
LITERATURE CITED 273

INTRODUCTION

What occasions a result? What is its determining cause?

We have an answer to questions of this sort in many specific cases, but none of the attempts to produce a general formula universally applicable for the solution of such questions has been entirely satisfactory. The present paper is a critical discussion of the latest solution offered, the method of "path coefficients" as proposed by WRIGHT (1921 a).

The conscious attempts to obtain a mathematical measure of causation, or to establish a mathematical criterion by which to test the truth of the statement that one event is the cause of another, have been, in the main, recent developments, but they are all essentially refinements of the simple method of concluding, because the observer has never known one event to happen without being followed by the other, that the two are therefore cause and effect. Although there may possibly be a few cases that appear to be exceptions when the observer is forming his conclusions, he is apt to reject them as being due to certain factors which he overlooked. This whole procedure is simply a non-mathematical way of determining roughly the degree of association or of correlation between the two events and regarding a high correlation either as causation itself or as evidence of a

[1] Papers from the Department of Biometry and Vital Statistics, School of Hygiene and Public Health, JOHNS HOPKINS UNIVERSITY, No. 42.

"causal relation." GALTON (1889) says in regard to correlation between organs:

"It is easy to see that correlation must be the consequence of the variations of the two organs being partly due to common causes. If they were wholly due to common causes, the co-relation would be nil. Between these two extremes are an endless number of intermediate cases."

This is the opinion of the man who appears to have been the first to conceive the idea of mathematical correlation (PEARSON 1920). The works of BRAVAIS and of GAUSS are treatments of the probability of errors of observation, and afford no basis for a claim that either of them discovered correlation.

"Causation" has been popularly used to express the condition of association, when applied to natural phenomena. There is no philosophical basis for giving it a wider meaning than partial or absolute association. In no case has it been proved that there is an inherent necessity in the laws of nature. Causation is correlation.

In his "Grammar of Science," PEARSON (1900), who developed the product-sum correlation coefficient now used, says in regard to scientific law and causation:

"Law in the scientific sense only describes in mental shorthand the sequences of our perceptions. It does not explain *why* those perceptions have a certain order, nor *why* that order repeats itself; the law discovered by science introduces no element of necessity into the consequences of our sense impressions; it merely gives a concise statement of *how* changes are taking place. That a certain sequence has occurred and recurred in the past is a matter of experience to which we give expression in the concept *causation*; that it will continue to occur in the future is a matter of belief to which we give expression in *probability*. Science in no case can demonstrate any inherent necessity in a sequence, nor prove with absolute certainty that it must be repeated."

. . . . . .

"When we say that we have reached a 'mechanical explanation' of any phenomenon, we only mean that we have described in the concise language of mechanics a certain routine of perceptions. We are neither able to explain why sense-impressions have a definite sequence, nor to assert that there is really an element of necessity in the phenomenon. Regarded from this standpoint, the laws of mechanics are seen to be essentially an intellectual product, and it appears absolutely unreasonable to contrast the mechanical with the intellectual when once these words are grasped in their accurate scientific sense."

. . . . . .

"No phenomenon or stage in a sequence has only one cause, all antecedent stages are successive causes, and, as science has no reason to infer a first cause, the succession of causes can be carried back to the limit of existing knowledge and beyond that *ad infinitum* in the field of conceivable knowledge. When we scientifically state causes we are really describing the successive

stages of a routine of experience. 'Causation' says JOHN STUART MILL 'is uniform antecedence' and this definition is perfectly in accord with the scientific concept."

. . . . . .

"The causes of any *individual* thing thus widen out into the unmanageable history of the universe. The causes of any finite portions of the universe lead us irresistibly to the history of the universe as a whole."

The above quotations are made, not as an appeal to authority, but because Professor PEARSON has already inimitably summarized the subject.

The theory of planetary motion is an intellectual concept that has been built up to describe, at least approximately, the observed events of the movements of the planets. If the theory holds, certain things *must* of logical necessity be true, but we must beware lest we unconsciously and illogically think that the necessity that lies in the concept also inheres in the order of nature which the concept attempts to describe.

The idea of determination, in the sense of causes fixing beforehand the nature of the effect, is based upon the belief in an inherent necessity in the order of things. We have seen that no such necessity can be proved. Therefore, determination should be used only in the sense of an ability to predict with fair accuracy the value of an effect when the values of its principal causes are known. This ability is based upon our knowledge of the degree of association between the causes and the effect.

To contrast "causation" and "correlation" is unwarranted because causation is simply perfect correlation. Incomplete correlation denotes partial causation, the effect here being brought about by more than one important cause. Many things show either high or perfect correlation that, on common-sense grounds, can not possibly be cause and effect. But we can not tell *a priori* what things are cause and effect and a conflict between our observations and our "common-sense" belief may be due either to an unwarranted belief or else to the calculation of our coefficients of correlation from too few data.

If Röntgen rays be directed against a brick wall, one can see through it. But it would indeed be difficult to imagine two more dissimilar things than Röntgen rays and sight through a brick wall; and yet, because these are invariably correlated, they are now so accepted. An example of high correlation and no causal relation might be the correlation over a two year period between the weight of a child born in 1917 and the tonnage production of ships in the United States. Here the data are evidently insufficient. We know that children born in 1917 grew at practically the same rate as children have always grown, but that ships were produced at a much faster rate in order to meet the war needs. Therefore if we correlate

the weight of children for their first two years of life with the tonnage production of ships over any long period we may be sure that the correlation coefficient would be practically zero.

It seems clear that perfect correlation, *when based upon sufficient experience*, is causation in the scientific sense.

THE METHOD OF PATH COEFFICIENTS

This method is claimed by WRIGHT (1921 a, b) to provide a measure of the influence of each cause upon the effect. Not only does it enable one to determine the effects of different systems of breeding, but provides a solution to the important problem of the relative influence of heredity and environment. To find flaws in a method that would be of such great value to science if only it were valid is certainly disappointing. The basic fallacy of the method appears to be the assumption that it is possible to set up *a priori* a comparatively simple graphic system which will truly represent the lines of action of several variables upon each other, and upon a common result.

In his introduction WRIGHT (1921 a) states:

> "The method depends on the combination of knowledge of the degrees of correlation among variables in a system with such knowledge as may be possessed of the causal relations. In cases in which the causal relations are uncertain the method can be used to find the logical consequences of any particular hypothesis in regard to them."

We have to set up a graphic system of the way we think the variables act upon each other. If we have enough observed correlation coefficients, we calculate the path coefficients and the coefficients of determination. The results are then compared with what we expect or have observed to be true in nature, and if they are in pretty close agreement our hypothesis is accepted, and we are to regard the system as showing the true relations between variables. If they are not in agreement, the hypothesis upon which we built our system must be wrong and a new one will have to be tried. But even if the observed and calculated values of the correlation coefficients agree, we can by no means be sure that we have set up the true system. An infinitude of values of x and y satisfy the simple equation $x^2+y^2=1$. The arrangement of the system depends entirely upon the judgment of the observer, and no *test* of that judgment follows in the least.

In all set-ups, or diagrams of systems, it is necessary to cut off the lines of causation at points not very far back in the chain of causes. This leaves two or more cause groups with nothing behind them, although it is

inconceivable that these groups have not some common causes. If we put into our system all important causes we know of, and all the important causes of these, and so on back, we would cover the whole universe and even then find no logical stopping place. There is absolutely nothing in the method to tell us how far back we should go. Apparently WRIGHT himself goes back as far as the observed correlation coefficients, which are needed to solve the equations, will permit. But extension backwards will change the values of path coefficients and coefficients of determination, and may also render the whole system unsolvable.

Two methods for the solution of the hypothetical systems are given; the direct, using determinants; and the indirect, using simultaneous equations. The indirect method is said to be less laborious than the direct, and this method "is more flexible in that it can be used to test out the consequences of any assumed relation among factors." (WRIGHT 1921 a, p. 578). Also (loc. cit., page 579):

> "One should not attempt to apply in general a causal interpretation to solutions by the direct methods. In these cases, determination can usually be used only in the sense in which it can be said that knowledge of the effect determines the probable value of the cause. This is the sense in which PEARSON's formula for multiple regression must be interpreted."

Measures of association or correlation are provided by mathematics, but we have no mathematical test which will enable us to tell absolutely whether or not to interpret any particular case as one of causation. We can not be sure that we have taken enough cases or a sufficiently large number of variables into consideration. By using the correlation coefficient we *know* that, in the sample of the universe which we have studied, certain variables were more or less closely associated than others, as indicated by the value of r, the coefficient of correlation, provided that the variations of each variable followed the Gaussian, or normal, curve of error. From this knowledge we are led to *believe*, either that the sample tried is not a fair one and another one is needed, or that certain causal relations probably do or do not exist. Statistical methods, particularly multiple correlation, indicate causes when they are used with common sense and upon the data of critical experiments. But the method of path coefficients does not aid us because of the following three fallacies that appear to vitiate this theory. These are (1) the assumption that a correct system of the action of the variables upon each other can be set up from *a priori* knowledge; (2) the idea that causation implies an inherently necessary connection between things, or that in some other way it differs from correlation; (3) the necessity of breaking off the chain of causes at some

comparatively near finite point. The applications of this theory in the latter part of this paper give impossible results and illustrate faults in the method.

THE MATHEMATICS OF PATH COEFFICIENTS

The section on Definitions in WRIGHT's paper opens with the following sentences (italics mine):

"We will start with the assumption that the direct influence along a given path can be measured by the standard deviation remaining in the effect after all other possible paths of influence are eliminated, *while variation of the causes back of the given path is kept as great as ever, regardless of their relations to the other variables which have been made constant.* Let X be the dependent variable or effect and A the independent variable or cause. The expression $\sigma_{X.A}$ will be used for the standard deviation of X, which is found under the foregoing conditions, and may be read as the standard deviation of X due to A."

If X is regarded as being completely determined by A and B, WRIGHT's $\sigma_{X.A}$ is somewhat like YULE's $\sigma_{X.B}$, the standard deviation of X when B is held constant; and when X is completely determined by A, B and C, it is somewhat like YULE's $\sigma_{X.BC}$. The physical interpretation of $\sigma_{X.B}$ and $\sigma_{X.BC}$ is very easy. If from a large group, all the cases having the same-size B's, or the same-size B's and C's, were picked and the standard deviation of the X's in this new group was found, the result would be the familiar $\sigma_{X.B}$ or $\sigma_{X.BC}$. But to make this correspond to WRIGHT's $\sigma_{X.A}$ we should in some manner have to keep all the variables back of B, which affected A, *just as variable as before.* How this might be done is difficult for the mind to grasp.

In figure 1, if we wish to get WRIGHT's $\sigma_{X.A}$ we must not let any action come along the path BX but must make as much as before come along $A\,X$, $A\,D$ and $A\,C$. If A and B are correlated, and if B is given a constant value it follows that the variation in C must be reduced and the path coefficient along the line $A\,C$ changed. In holding constant a variable we are really picking out only observations of it that have the same value and are considering the causes and effects of this new group. The results of a constant effect are obviously less variable than those of a variable effect; and, although there may conceivably be some compensatory changes in the causes, it seems impossible that they should vary as much in producing a constant effect as in producing a variable one. This simple point of logic WRIGHT appears to have overlooked.

In a paper in Genetics, WRIGHT (1921 b) says:

"A path coefficient differs from a coefficient of correlation in having direction. The correlation between two variables can be shown to equal the sum of the products of the chains of path coefficients along all the paths by which the variables are connected."

The pure mathematics by which this is shown is apparently faultless in the sense of mere algebraic manipulation, but it is based upon assumptions which are wholly without warrant from the standpoint of concrete, phenomenal actuality.

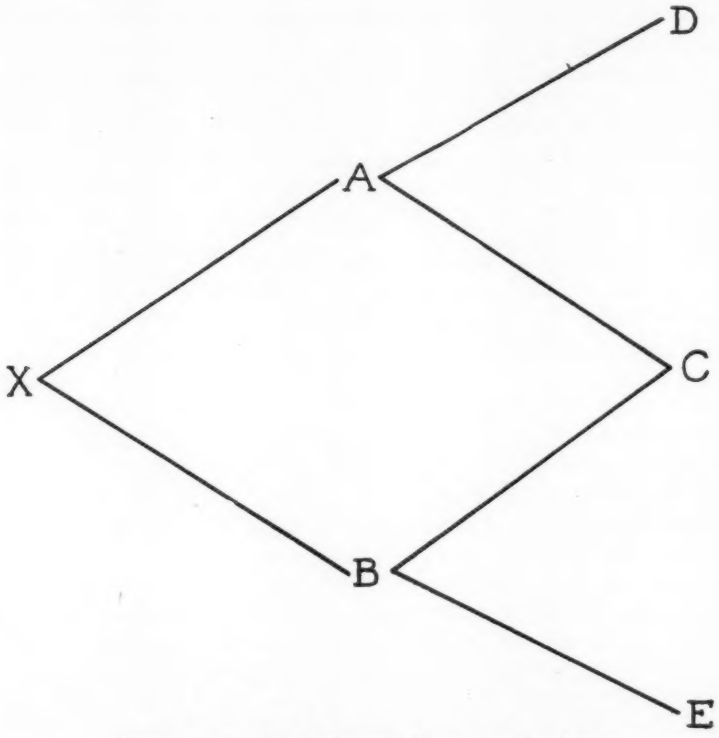

FIGURE 1.—An effect, X, determined by two correlated causes.

WRIGHT'S GUINEA-PIG EXAMPLE

The guinea-pig is "intended merely to furnish a simple illustration of the method" (WRIGHT 1921 a, p. 570). It shows us how to measure the relative importance in determining the birth weight of guinea-pigs, X, of the factors Q, prenatal growth curve; P, gestation period; L, size of litter; A, heredity and environmental factors which determine Q apart from size of litter; C, factors determining gestation period apart from size

of litter. The "prenatal growth curve" is apparently the average growth per unit of time during the gestation period. If we multiply the average growth by the time we necessarily get the birth weight, but it is impossible to get the average growth until the total growth and the time are known. If $\frac{X}{P}=Q$, then any two of the variables mathematically determine the third. It is as logical to say that the birth weight and the gestation period determine the prenatal growth curve, as to say that the gestation period and the prenatal growth curve determine the birth weight.

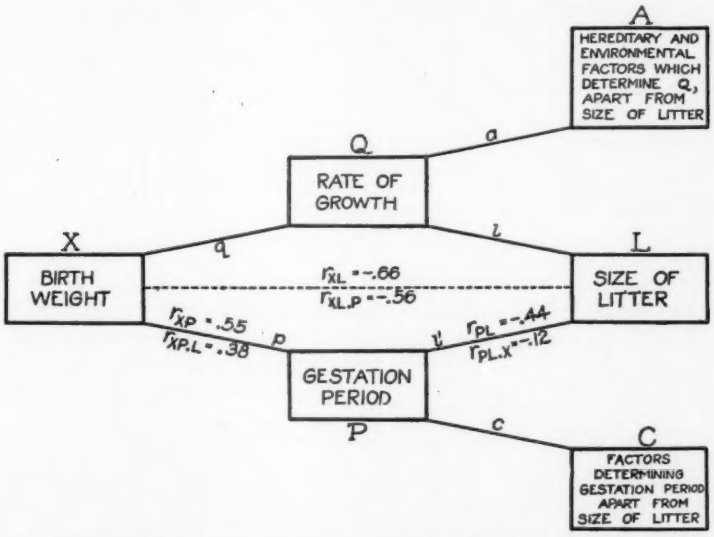

FIGURE 2.—The system set up by WRIGHT (1921 a) for finding the relative importance of factors determining the birth weight of guinea-pigs.

In solving the guinea-pig problem three things are known from experience (figure 2). These are the correlations between birth weight and interval between litters, which is assumed to be the gestation period if less than 75 days, $r_{XP}=+.5547$; birth weight and size of litter, $r_{XL}=-.6578$; and between gestation period and size of litter, $r_{PL}=.4444$. These are the realities. From the general equation where the coefficient of correlation is the sum of the products of the path coefficients along all the chains of causes connecting the two variables, he derives three equations:

(1) $r_{XP}=p+q\,l\,l'$
(2) $r_{XL}=q\,l+p\,l'$
(3) $r_{PL}=l'$

We might get equations (1) and (2) directly from figure 2, but to be consistent throughout we must get equation (1) from the system shown in

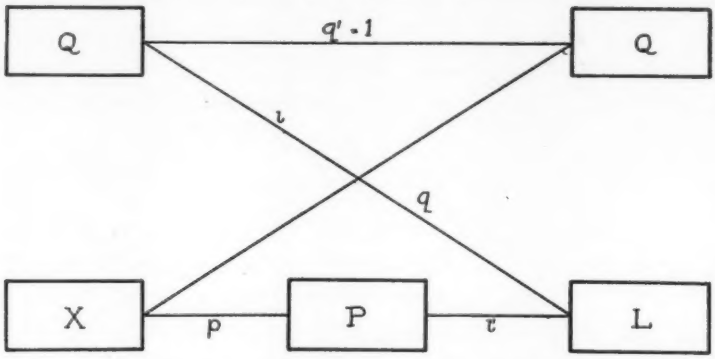

FIGURE 3.—System from which equation (1) would be obtained.

figure 3, equation (2) from figure 4, and equation (3) from figure 5. In each of these systems one of the variables is made a cause of itself with a path coefficient of unity between variable as cause and variable as effect.

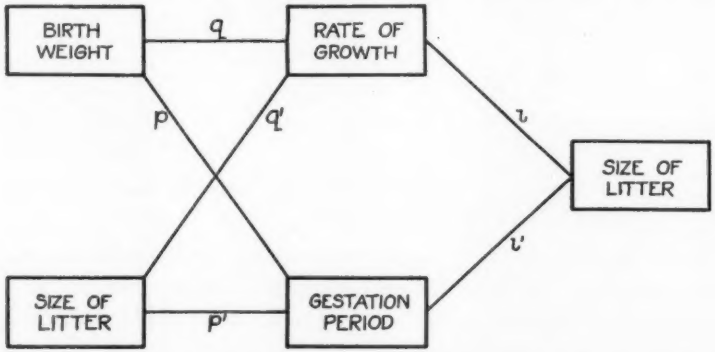

FIGURE 4.—System from which equation (2) would be obtained. $q'l=1$, and $p'l'=1$.

This seems rather forced, but if we attempt to obtain the relation from the original diagram, what is there to prevent our setting $r_{PL}=l'+pq\,l$; that is, following all the possible paths of the original set-up in getting

equation (3)? Perhaps in tracing the chains of causes we are not allowed to come to the effect through something that follows it. Such a rule would be meaningless when the true relation of cause and effect, that of invariable association, is kept in mind.

Three more equations are based upon the fact that the sum of the coefficients of determination of any effect must be equal to unity. These coefficients are simply the squares of the path coefficients between cause

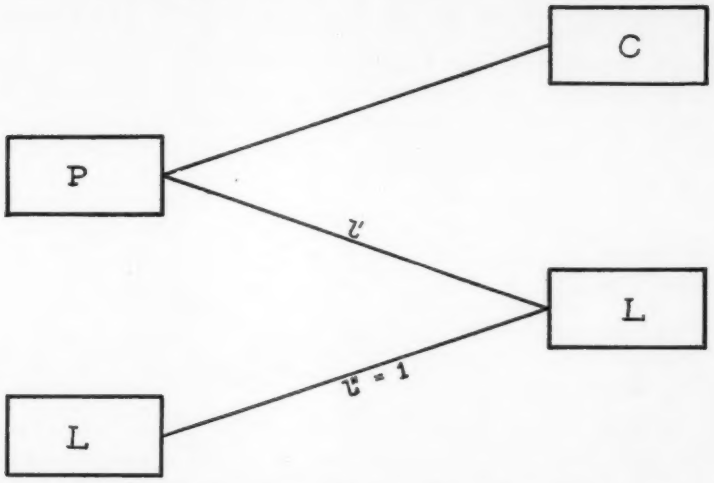

FIGURE 5.—System from which equation (3) would be obtained.

and effect; except when there are correlated causes, when there is also a coefficient of determination which represents the action of the correlated causes taken together. This coefficient is twice the product of the path coefficients from each cause to the effect, times the coefficient of correlation between the causes. The additional equations in this case are

$$(4)\quad q^2+p^2+2q\,p\,l\,l'=1$$
$$(5)\quad a^2+l^2=1$$
$$(6)\quad l'^2+c^2=1$$

The values obtained from the six equations are assumed to be measures of realities if the diagrams accurately represent the causal relations. Figure 6 shows the values obtained for each path coefficient. No value of the probable error of any constant is given by WRIGHT. If the method of path coefficients were valid, a knowledge of the probable error of any constant

would be essential in many cases. The correlation between size of litter and gestation period for constant birth weight, using only the observed r's the writer computed and found to be $r_{PL.X} = -.12$. How are we to account for the difference between this value and $r_{PL} = l' = -.44$? The $r_{PL.X}$ means that when we select groups of guinea-pigs of the same birth weight the correlation between size of litter and gestation period is greatly reduced, and the path coefficient and coefficient of determination are correspondingly reduced, the latter taking the value $-.014$. Therefore,

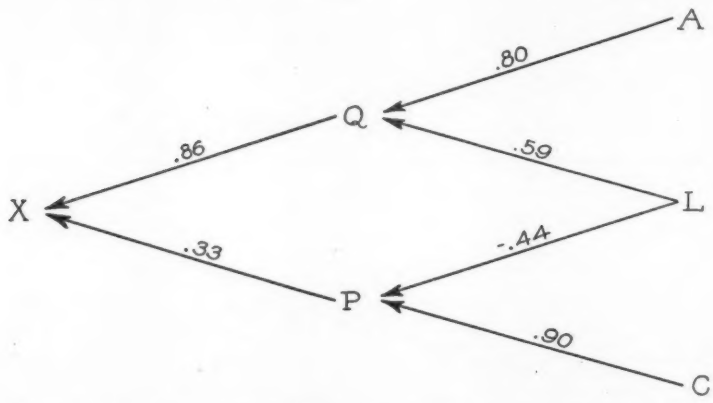

FIGURE 6.—Showing the values obtained by WRIGHT (1921 a) for the path coefficients between the factors determining birth weight of guinea-pigs. See figure 2.

when guinea-pigs of equal birth weight are considered, the size of the litter has practically no effect upon the gestation period.

TEST OF WRIGHT'S METHOD

Except in unusual cases we can check the results of this method of WRIGHT'S only by testing them with what we *think* on common sense grounds ought to be true. In the hands of a man well acquainted with the realities in the field he is investigating, this method would be likely to lead to results not far from the truth, because if any values appear to be inconsistent, a new set-up of causes and effects will be made. Guesses by a trained man would be on the whole quite as good and much less work; whereas an untrained man can not be sure of the validity of his results at all, because he is not familiar with the realities in the field of study.

Let us attempt to apply this method to two examples where we know more of the correlation coefficients, and can get our path coefficients and coefficients of determination in more than one way.

Example 1

We are interested here in the relative part played by the number of seeds per pod, and the number of ovules per pod, in determining the seed weight of the seeds produced. We set up the diagram shown in figure 7, making (1) seed weight, be determined by (2) seeds per pod, (3) ovules per pod, and (4) other causes than (2) and (3). Seeds per pod is determined by (5) pods per plant and (6) other causes than (5) which affect (2). Ovules per pod is also determined by (5) and by (7) a group of other causes. Behind these we do not go as we have not enough observations

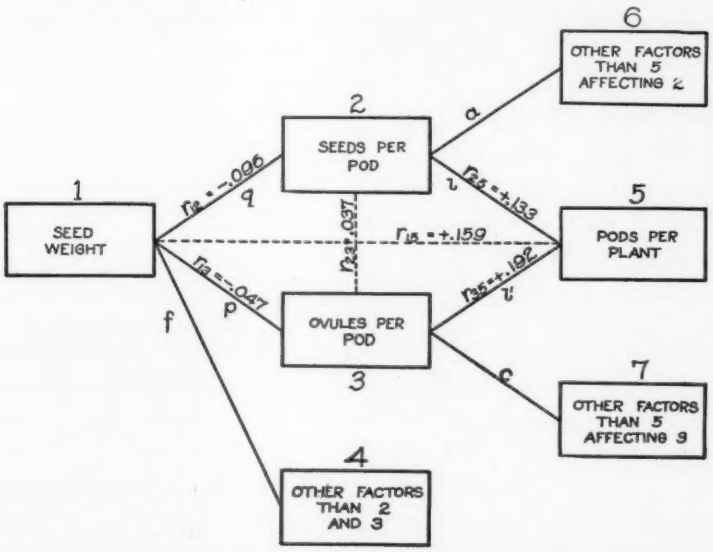

FIGURE 7.—The system set up for example 1, showing the observed correlation coefficients.

to solve a more complicated system and we will assume that we are to be satisfied with approximate results. This figure is identical with the diagram (figure 2) used by WRIGHT (1921 a) in setting up his equations, except that we have an all-other-causes path affecting (1) directly. This enters only in the equation which makes the summation of the coefficients of de-

termination equal to unity. The correlation coefficients shown in the figure are from J. ARTHUR HARRIS (1913 a, 1913 b, 1916) and although of low absolute values, they are very probably significant, because the probable errors where given are extremely small, and all the constants appear to have been based upon a very wide experience.

The equations involving only the first powers of the path coefficients are:

$$(1)\ r_{13}=p+q\,l\,l'$$
$$(1\text{ bis})\ r_{12}=q+p\,l\,l'$$
$$(2)\ r_{15}=q\,l+p\,l'$$
$$(3)\ r_{25}=l$$
$$(4)\ r_{35}=l'$$

Taking r_{13}, r_{15}, r_{25}, and r_{35} as known, we will use the above equations to get the path coefficients and the coefficients of determination.

Substituting the numerical values of r_{13}, l and l' in equation (1) we have

$$p=-.047-(.133\times.192)\,q$$
$$=-.047-.025536\,q.$$

Equation (2) now becomes

$$.159=.133\,q+.192\,(-.047-.025536\,q)$$
$$q=1.3117.$$

Whence

$$p=-.08049$$

These values give in (1bis)

$$r_{12}=1.312-(.080\times.133\times.192)$$
$$=+1.310$$

Assuming r_{12}, r_{15}, r_{25}, and r_{35} known and solving for r_{13} we get

$$q=-.119$$
$$p=+.911$$

These values in (1) give

$$r_{13}=+.908$$

As a correlation coefficient can never be greater than 1, $r=1.310$ is impossible. The computed values of r_{12} and of r_{13} are in both cases more than twelve times the observed values and opposite to them in sign. Such results are ridiculous.

Let us now test this system for the coefficient of determination of the causes which we considered as "all other causes," group 7. From the principle that the sum of the coefficients of determination must be unity, we have

$$(5)\ q^2+p^2+2\,qpll'+f^2=1$$

Substituting in this the values obtained for p and q in our first solution, we have

$$1.7213+.0064-.0054+f^2=1$$
$$f^2=-.7223$$

Substituting the values from the second solution, we have

$$f^2=.1615$$

What does this really tell us about the effects of the factors not included? In the first case the determination is negative and has no meaning. The proportion of the standard deviation of the seed weight due to all other causes than those acting through seeds per pod and ovules per pod is $\sqrt{-.7223}=f$ according to WRIGHT's theory. This is an imaginary standard deviation, a thing not encountered in statistics. In the second case we find that the unknown causes play apparently a real although small part in determining the seed weight. But there is no inherent or logical reason *in* WRIGHT's *theory* why the first solution is not as good as the second.

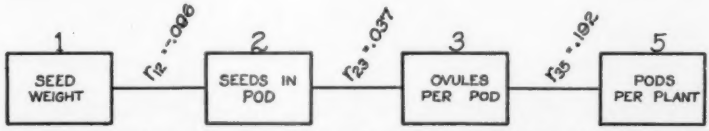

FIGURE 8.—An alternative set-up for the seed-weight example, giving observed correlation coefficients.

WRIGHT gives a special formula for finding the coefficient of determination of factors not specifically included in a system when correlations between the factors included are known. In our case the appropriate formula seems to be the one for two known correlated causes acting upon the effect. Using this formula and the observed r's gives the coefficient of determination between not-included causes and the effect, or f^2, equal to 0.9944, but using the calculated r's gives f equal $\sqrt{-1.482}$. Here is another case of two widely different values for the same constant, and one value is again imaginary and impossible.

It may be contended that the causal connections are really a straight line from pods per plant to ovules per pod, to seeds per pod, to seed weight. In this case we draw our diagram as in figure 8. If we multiply together the path coefficients from 1 to 5, we should obtain the correlation between 1 and 5 because there are no common causes and the path coefficients therefore are equal, on WRIGHT's theory, to the coefficients of the correlation. By multiplying we find $r_{15}=-.000682$. This is not even one percent of the observed value. Evidently the theory works with this set-up no better than with the last.

Example 2

This is an application of the method of path coefficients in an attempt to determine approximately the relative importance of some factors influencing the amount of heat produced in human basal metabolism. As shown in figure 9, we assume that stature determines in part body weight and body surface, and that these, with a group of other factors, determine the heat produced. The correlation coefficients given in the figure are taken directly from HARRIS and BENEDICT (1919) with the exception of the one between surface and stature which had to be computed from the raw data given in the reference. This figure is identical in form with figure 7

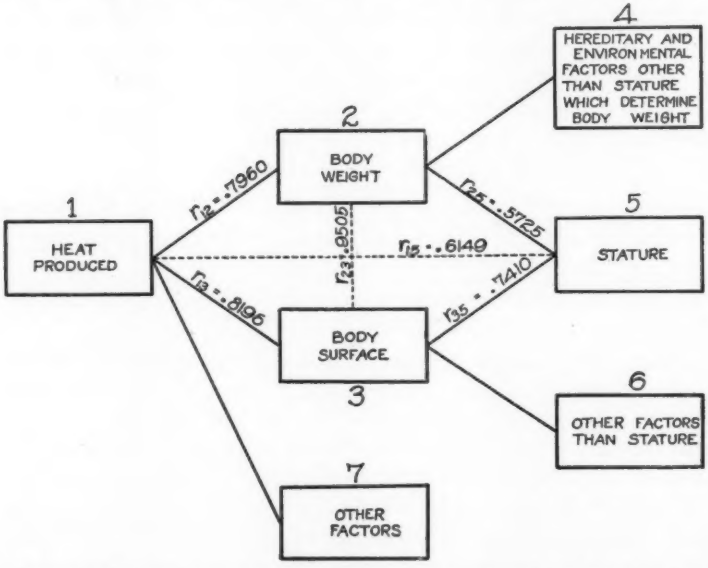

FIGURE 9.—The system set up for example 2, showing the observed correlation coefficients.

used in example 1, and the same set of equations is therefore applicable. Without repeating these equations we will give the results obtained by solving them. Treating r_{12} as unknown we find its value to be +.3718, or less than one-half of the observed value. Treating r_{13} as unknown we find its value to be +.5997 or about three-fourths of the observed value. The correlation between the heat produced, and factors other than body weight and surface and those acting through them, is found to be either (a) r_{10} = +.5703, or (b) $\sqrt{-1.3893}$, or (c) +.338, depending upon the

values used for r_{12}, and r_{13}. The values of the path coefficients are found to be $p = +.8072$ or $+.3196$, and $q = +.0293$ or $+.6604$. The results of this example, like those of the preceding, are inconsistent with themselves and with reality. When we see that the path coefficients are unreliable in these cases where we can check them, we are not likely to place any great confidence in them where they cannot be checked.

Should the criticism be made that these examples give absurd results because the true action of the variables upon each other is not truly represented in the diagram the writer would reply, first, that such criticism but strengthens one of his main points; namely, that it is impossible to tell *a priori* how the system should be set up, and that the closeness of agreement between calculated or expected and observed values is an unscientific criterion by which to judge the validity of such a system; and second he would invite careful examination from a biological standpoint of his diagrams and WRIGHT'S, with a view of the reader's seeing for himself whether the one set is more unfair or less related to the probable truth than the other.

CONCLUSION

We therefore conclude that philosophically the basis of the method of path coefficients is faulty, while practically the results of applying it where it can be checked prove it to be wholly unreliable.

The writer believes himself still open-minded on WRIGHT'S proposition, but has an even more intense conviction that before that author's contribution to the theory of partial correlation can be taken seriously he will have to bring forward evidence altogether more cogent in respect of both logic and fact than any he has so far adduced.

LITERATURE CITED

GALTON, FRANCIS, 1889 Co-relations and their measurement. Proc. Roy. Soc. **45**: 135–145.

HARRIS, J. ARTHUR, 1913 a On the relationships between the number of ovules formed and the capacity of the ovary for maturing its ovules into seeds. Bull. Torrey Bot. Club. **40**: 447–455.

1913 b A quantitative study of the factors influencing the weight of the bean seed. 1. Intraovarial correlation. Beih. Bot. Centralbl. **31**: 1–12.

1916 A quantitative study of the factors influencing the weight of the bean seed. 2. Correlation between number of pods per plant and seed weight. Bull. Torrey Bot. Club **43**: 485–494.

HARRIS, J. ARTHUR, and BENEDICT, F. G., 1919 A biometric study of basal metabolism in man. Carnegie Inst. Washington Publ. 279. 266 pp.

PEARSON, KARL, 1900 Grammar of science. 590 pp. London: Adam and Charles Black.

PEARSON, KARL, 1920 Notes on the history of correlation. Biometrika **13**: 25-45.

WRIGHT, SEWALL, 1921 a Correlation and causation. Jour. Agric. Res. **20**: 557–585.

1921 b Systems of mating. 1. The biometric relations between parent and offspring. Genetics **6**: 111–123.

THE CORRELATION BETWEEN FIRST- AND SECOND-YEAR EGG PRODUCTION IN THE DOMESTIC FOWL

J. ARTHUR HARRIS

Station for Experimental Evolution, Cold Spring Harbor, Long Island, New York

and

H. R. LEWIS

New Jersey Agricultural Experiment Station, New Brunswick, New Jersey

Received August 6, 1921

TABLE OF CONTENTS

INTRODUCTION 274
Historical 275
Materials and methods 277
Analysis of data 278
Inter-annual correlation for total production in the flock as a whole 278
Inter-annual correlation within the individual pens of the contest flock 280
Inter-annual correlation of flock means 283
The correlation between the monthly records of one year and the annual record of another year 284
Inter-annual correlation for monthly production 287
Genetic significance of the inter-annual correlation for monthly production 293
Relationship between the condition of the birds with respect to laying activity during individual months of the first year and the annual record of the second year, with special reference to culling for second-year production 302
DISCUSSION AND CONCLUSIONS 313
LITERATURE CITED 317

INTRODUCTION

It seems rather remarkable that the problem of the relationship between the egg production of the first and second year in the domestic fowl should have received no wider attention at the hands of practical poultrymen or experimental investigators. As far as we are aware the only quantitative discussions of the subject have been those of PEARL and SURFACE (1909), of NIXON (1912), and of BALL, ALDER and EGBERT (1916).

From the theoretical side the interrelationships between the physiological activities of the different periods of the life of the organism is a problem of independent importance. It is, furthermore, impossible to solve the theoretical problem of the mode of inheritance of fecundity in

the fowl without detailed knowledge of the distribution of fecundity and of the interrelationships between the fecundity of diverse periods.

From the practical side it will be clear at once that if there be a marked correlation between the record of the first- and the second-year's production of the individual it will be possible to extend the method of prediction of future egg production, now being developed for the first year (HARRIS, KIRKPATRICK and BLAKESLEE 1921; HARRIS, KIRKPATRICK, BLAKESLEE, WARNER and CARD 1921), to second-year production. Such methods of prediction will be practicable[1] in the case of trap-nested birds or of birds selected by the color criteria first investigated in a detailed statistical manner by one of us in coöperation with BLAKESLEE, WARNER and KIRKPATRICK (HARRIS, BLAKESLEE and WARNER 1917; HARRIS, BLAKESLEE, WARNER and KIRKPATRICK 1917; BLAKESLEE, HARRIS, WARNER and KIRKPATRICK 1917).

It is possible, therefore, that a knowledge of either of these criteria will make possible the extension of the method of culling, with a resulting material increase of the second-year production.

In relation to the problem of commercial egg production, it may be noted that the bird's second-year production may in commercial work be a factor not to be ignored. Those who have dealt with the problem of the inheritance of fertility from the experimental side have in general tacitly assumed that only first-year records require consideration. But this has not been the case invariably; BALL, TURPIN and ALDER (1914), BALL, ALDER and EGBERT (1916) and BALL and ALDER (1917) fully recognized the desirability of obtaining more than the first-year record of the bird as a measure of the egg-producing capacity of the organism. BALL, ALDER and EGBERT (1916, p. 4) write:

> ". . . It seemed necessary, before intelligent selection could be practised to discover:
>
> First, a reliable measure of the productivity of a hen.
>
> Second, the average productive life of a hen (that should be considered in selection). With these two points established it was thought that it would then be possible to plan an experiment to determine the value of selection in improving a strain of birds."

HISTORICAL

GOWELL as early as 1903 recognized the importance of the problem of inter-annual correlation when he wrote (GOWELL 1903, p. 74):

[1] While they may be scientifically practicable they may not be economically feasible in all commercial plants.

"With many poultrymen the idea is prevalent that if a hen lay but few eggs the first year, she is likely to do better the second year, than though she had laid well during the first.

The data so far secured does not show that hens that yield 120 eggs, or less, the first year, yield satisfactorily the second year. Those that yielded in the vicinity of 100 the first year have yielded very lightly the second year."

As far as we are aware the first determination of the relationship between the production of the first and second year of the life of the bird is that based by PEARL and SURFACE (1909, pp. 63-66) upon the data for the 66 Barred Plymouth Rock birds recorded by GOWELL (1903, p. 75). They state the correlation to be

$$r = +0.032 \pm 0.083$$

and conclude:

"While this conclusion is positive, yet it is extremely small and obviously insignificant in comparison with its probable error. We can only conclude that so far as any evidence from our present material is concerned there is no sensible correlation between the egg production of the first laying year and that of the second."

PEARL (1912, p. 169) says:

". . . the potential 'anatomical' fecundity is very much higher than the actually realized fecundity. This is true even if we suppose the bird to be allowed to live until it dies a natural death. *Experience shows that birds which make a high fecundity record in the first year of their life, generally speaking, never do so thereafter.*"[2]

While this conclusion is at variance with the statements made by GOWELL in recording the original data, and is in itself demonstrably incorrect,[3] it seems to have been widely accepted by geneticists and has probably had a serious deterrent influence on the investigation of the economically highly important problem of the interrelationship of first- and second-year egg production. It has also tended to focus attention upon first-year record only in the investigation of the problem of inheritance of fecundity in the fowl. These problems will be discussed in detail later.

NIXON in 1910 clearly recognized (NIXON 1912) the essential problems of inter-annual correlation in egg production when she stated that the purpose of her study of records of 88 White Leghorns was

". . . . to show, first, whether hens which produce well for any particular year are likely to give good egg production for a longer period; second, whether it is possible to judge from the egg record of any particular year what is likely to be the egg production of the same hens during succeeding years."

[2] Not italicized in original.

[3] The decimal point has been misplaced at some step in the calculation of the coefficient. The correct value is $r = .327 \pm .974$, a constant in excellent agreement with many of those given later in the present paper.

The 88 birds considered gave a low average production[4] in each of the three years. The averages were 92.4, 96.9, and 86.2 eggs in the three years. Using a class interval of 25 eggs she deduced the following correlations between the production of the different years:

First and second, $r = 0.548 \pm 0.050$
First and third, $r = 0.153 \pm 0.070$
Second and third, $r = 0.397 \pm 0.061$

She also considers the relationship between the production of the three individual years and the production of the entire period. These results do not, however, concern us in this place.

The most extensive series of studies of the production of birds for a period of years is that carried out at the UTAH AGRICULTURAL EXPERIMENT STATION. In 1914 BALL, TURPIN and ALDER analyzed the records of a seven-year-old flock of White Leghorn hens and their progeny. In a later publication BALL, ALDER and EGBERT (1916) considered, among other problems, the relationship between the egg records of first and second, first and third, and second and third years. In this paper they bring together the literature on the egg-laying capacity of birds in years subsequent to the first. Finally BALL and ALDER (1917) treat the problem of the seasonal distribution of egg production.

The constants deduced from the Utah data are in full agreement with those presented later in this paper in that they indicate a positive correlation between the egg production of the individual in various years. There are, however, a number of points of theory which fall outside the scope of the data upon which this paper is based; in consequence it seems wisest to postpone a detailed discussion of these data until a comparative treatment of the various series of egg records now being analyzed is possible.

MATERIALS AND METHODS

The records analyzed represent the entire first- and second-year egg-laying performance of 443 White Leghorn birds submitted in the first VINELAND INTERNATIONAL EGG-LAYING AND BREEDING CONTEST. Originally 540 birds were entered but a number died during either the first or second year.

Details concerning the regulations governing the contest and the conditions under which the birds were maintained may be derived from previous publications (LEWIS 1917, 1918; LEWIS, HANNAS and WENE 1920).

[4] The records were derived from six flocks. In many cases these were under experimental treatment, and fed different rations or in a different manner.

The methods of statistical analysis are those in common use in modern higher statistics. Details are given where necessary.

ANALYSIS OF DATA

Inter-annual correlation for total production in the flock as a whole

We will first of all consider the correlation between the total annual production of the first and of the second year of the bird. The constants are the following:

First year	*Second year*
Mean = 174.05 ± 1.29	Mean = 139.79 ± 1.19
σ = 40.10 ± 0.91	σ = 37.05 ± 0.84
C.V. = 23.04 ± 0.55	C.V. = 26.50 ± 0.64

$r = .5485 \pm 0.0224$

This coefficient shows a medium degree of correlation between the first- and second-year production of the birds.

TABLE 1

Mean second-year production of birds making various records in first year and mean first-year production of birds making various records in second year.

RECORD OF FIRST YEAR	SECOND YEAR		RECORD OF SECOND YEAR	FIRST YEAR	
	f	Mean record		*f*	Mean record
0– 14	1	0.00	0–14	4	123.75
15– 29	—	———	15– 29	—	———
30– 44	—	———	30– 44	1	112.00
45– 59	—	———	45– 59	3	154.67
60– 74	3	110.00	60– 74	6	119.50
75– 89	5	90.60	75– 89	23	125.26
90–104	6	108.83	90–104	35	146.97
105–119	24	109.00	105–119	54	159.59
120–134	28	113.46	120–134	70	169.60
135–149	51	123.04	135–149	65	170.78
150–164	59	129.80	150–164	66	181.86
165–179	73	140.36	165–179	53	201.04
180–194	62	148.56	180–194	38	203.95
195–209	46	154.39	195–209	17	206.76
210–224	42	158.64	210–224	5	230.22
225–239	23	173.83	225–239	2	208.00
240–254	10	175.40	240–254	—	———
255–269	8	188.50	255–269	—	———
270–284	—	———	270–284	1	209.00
285–299	2	141.50	285–299	—	———

It is idle to attempt a discussion of the distribution of the mean egg production of the bird's second year associated with given records of the bird's first year without grouping the highly variable annual egg produc-

tions of the first year in classes. Adopting a class range of 15 eggs we have the frequencies and means of arrays shown in table 1.

Using grouped records for the first year and ungrouped records for the second year, we find the following correlation coefficient and regression equation.

$$r_{E_1E_2} = 0.5387 \pm 0.0227,\ E_2 = 53.046 + 0.499\ E_1$$

For the grouped values of the second year and the ungrouped values of first year production, the numerical values are

$$r_{E_1E_2} = 0.5541 \pm 0.0223,\ E_1 = 91.104 + 0.593\ E_2$$

The correlations calculated from the grouped frequencies are practically identical with those calculated from the ungrouped values. The

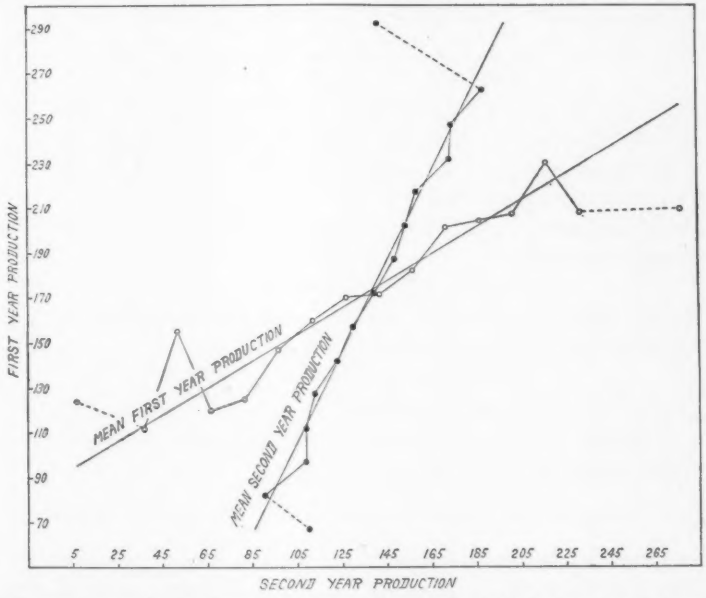

DIAGRAM 1.—Straight-line regression equations showing relationship between first- and second-year production.

constants have been recomputed with one of the two variables grouped, in order to have proper constants based on the same grouping as that employed in the regression diagrams.

The two equations are represented in diagram 1 which shows the mean egg production for the first year associated with various records in the second year and the mean production of the second year associated with

various records during the first year. The means are taken from table 1. Except for the higher and lower grades of production, the straight lines give excellent representations of the series of means.

For a formal test of linearity of regression we turn to the correlation ratio and BLAKEMAN'S test for linearity of regression.

The values of η (PEARSON 1905) calculated from the formula

$$\eta = \sigma_{\bar{E}}/\sigma_E$$

where $\sigma_{\bar{E}}$ is the standard deviation of the mean egg production of the arrays and σ_E is the standard deviation of the egg production of the flock as a whole, serves as a measure of the interrelationship of the two variables whether regression be linear or not. BLAKEMAN'S criterion (BLAKEMAN 1905)

$$\frac{\zeta}{E} = \frac{\sqrt{N}}{.67449} \quad .\tfrac{1}{2}\sqrt{\zeta}\,.\,\frac{1}{\sqrt{1+(1-\eta^2)-(1-r^2)^2}},$$

when $\zeta = \eta^2 - r^2$, furnishes, in connection with Miss GIBSON'S (1906) table, a convenient test for the linearity of the regression of the production of one year on the production of another year. The values of η and of BLAKEMAN'S criterion are as follows:

For the regression of second-year production (ungrouped) on first-year production (grouped),

$$\eta = 0.5593 \pm .0220, \quad \zeta/E_\zeta = 2.39$$

For the regression of the first-year production (ungrouped) on the second-year production (grouped),

$$\eta = 0.5815 \pm .0212, \quad \zeta/E_\zeta = 2.81$$

In both cases ζ/E_ζ is not far from 2.5 and we may reasonably assume that regression is, within the limits of the errors of random sampling, practically linear in both cases.

Inter-annual correlation within the individual pens of the contest flock

Two suggestions may be offered in explanation of the correlation between the egg record of the first and second year of the individual birds, demonstrated in the foregoing section. First, that the correlation between the first- and second-year record of the bird is due to her own individuality with respect to egg-laying capacity within the flock from which she is drawn. Second, that the several flocks from which the exhibit birds are selected differ widely in both first- and second-year egg record, and that as a consequence of the combination of the pens from the various flocks there is a correlation between the first- and second-year production of the individual birds in the contest flock as a whole.

To determine to what extent this correlation is due to the individuality of the birds and to what extent it is attributable to the differentiation of the flocks from which the birds were drawn we may (a) employ the method of fractional inter-class correlation coefficients developed elsewhere (HARRIS 1913), and (b) determine the inter-annual correlation coefficients for the birds of each individual pen.

TABLE 2

Mean annual production for two years, and correlation between annual production for two years, in individual pens.

PEN	BIRDS IN PEN	MEAN FOR FIRST YEAR	MEAN FOR SECOND YEAR	$r_{E_1E_2}$	PEN	BIRDS IN PEN	MEAN FOR FIRST YEAR	MEAN FOR SECOND YEAR	$r_{E_1E_2}$
41	10	192.3	143.4	+.7004	68	9	166.3	139.2	+.5097
42	7	205.7	172.4	+.8842	69	8	190.8	133.3	+.6899
43	9	200.7	157.1	+.6356	70	9	179.7	152.7	+.0949
44	10	141.7	111.9	+.7596	71	10	182.0	143.4	+.3610
45	8	177.0	142.5	+.9208	72	7	172.1	160.6	+.5911
46	6	190.3	154.2	+.7538	73	5	104.6	155.6	−.0552
47	9	166.2	134.0	+.8316	74	10	164.6	123.2	+.8699
48	7	166.6	149.6	+.6733	75	9	151.8	127.0	+.8807
49	8	169.4	134.9	+.7168	76	9	212.2	173.3	+.2123
50	8	175.6	150.6	+.9182	77	8	170.0	140.9	+.9348
51	8	174.1	105.3	+.4575	78	7	169.6	148.0	+.8953
52	8	161.6	123.1	+.0398	79	9	159.4	127.0	+.7047
53	8	167.5	141.8	+.4117	80	10	163.6	144.0	+.8568
54	5	185.4	149.4	−.2563	81	10	186.2	135.9	+.8295
55	8	182.4	129.1	+.5755	82	8	184.5	145.6	+.6490
56	8	131.1	118.0	+.4974	83	9	173.0	131.3	+.5403
57	9	165.6	134.8	−.1037	84	7	147.7	109.0	+.6581
58	9	158.6	117.7	+.6773	85	10	130.5	124.8	+.7862
59	5	148.4	127.0	+.4880	86	10	167.8	146.6	+.5183
60	7	194.6	118.7	+.4526	87	8	152.9	109.6	+.6495
61	9	140.7	146.1	+.8362	88	9	225.0	143.9	+.7467
62	9	210.9	148.2	+.2480	89	6	199.2	111.0	−.4805
63	7	144.6	122.1	+.8095	90	7	201.3	140.4	−.0090
64	8	182.4	157.9	+.7217	91	6	159.8	119.3	+.6699
65	9	185.1	172.2	+.7186	92	8	202.6	180.5	+.2398
66	7	187.1	148.9	+.5066	93	10	187.8	154.5	+.2037
67	9	178.8	166.0	−.1529	94	10	188.5	145.6	+.8379

Using formulae xvi–xviii of the paper cited, letting x = first-year and y = second-year productions, and remembering that the number of birds is the same in each of the two years, we find for the correlation between the first-year record of a bird and the second-year record of the $n-1$ other individual birds (other than herself) of the same pen

$$r = +.0786 \pm .0318, \; r/E_r = 2.5$$

The coefficient is positive in sign, and may be statistically significant. If it can be looked upon as statistically trustworthy it indicates that in the long run the first-year record of an individual bird furnishes some basis of prediction concerning the second-year record of other birds of the same pen. It is, however, very small as compared with the correlation, $r = .548$, for the two years of the individual bird's life.

This result proves that *it is the individuality of the bird within her own flock and not the individuality of the flocks which is the primary factor in determining the inter-annual correlation.*

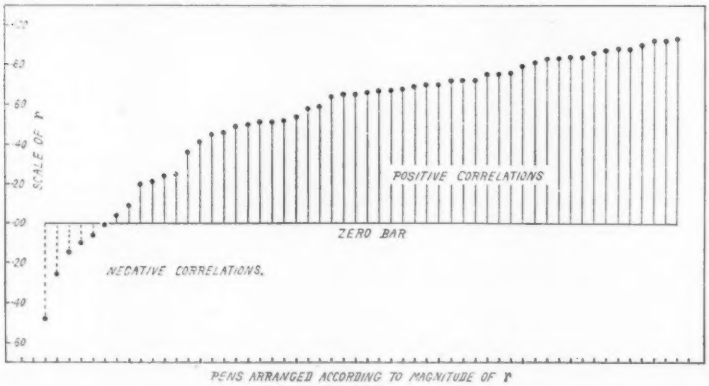

DIAGRAM 2.—Distribution of magnitudes of correlation coefficients showing the relationship between the first- and second-year record of birds of the same pen.

The problem may be approached from the other side by the determination of the inter-annual correlation within the individual pens.

While correlation coefficients based on series in which ten is the maximum number, can have little value individually, it has seemed desirable to determine the correlation between the first- and second-year production of the birds of the individual pens of the VINELAND CONTEST.

The correlations between the first- and second-year production within the individual flocks are shown in table 2. The first- and second-year means in this table will be used for another purpose later, and need not concern the reader for the moment. The coefficients are for the most part positive, and many of them are rather large. The distribution of the 54 coefficients, arranged in the order of their magnitude, are shown in diagram 2. A glance at the graph will show that only six pens show a

negative correlation, and that the magnitude of these coefficients is small as compared with the mass of the positive coefficients.[5]

The mean value of the correlation between the first- and second-year production of birds of the same pen is

$$r = .5890 \pm .0307,$$

which is very nearly the same as the $r = .5485$ demonstrated for the whole 443 birds available for the two years.

Inter-annual correlation of flock means

From the practical standpoint it is important to determine whether a flock which gives a high yield in one year may be expected to be characterized by high production in the second year.

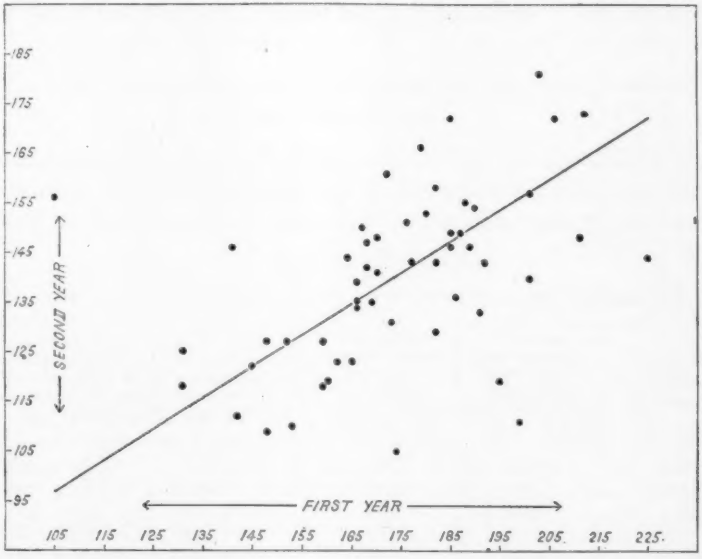

DIAGRAM 3.—Relationship between the average first- and second-year production of the pens.

Samples of ten birds or fewer are of course quite inadequate to represent the flocks from which they are drawn, and we must in consequence expect

[5] The average number of birds per pen for the 48 pens showing positive correlation is 8.4 whereas the average number for the 6 pens showing negative correlation is 6.8. The smaller number of birds in the pens showing negative correlation, as well as the small magnitude of these correlations, suggests that the negative sign of these coefficients is due merely to the probable errors of random sampling when very small samples are available, rather than that it represents a true biological relationship.

the correlation to be lower numerically than they would be found to be if based on means derived from larger numbers of birds.

Table 2 contains the pen averages for the two years and the number of individuals on which they are based.

The correlation between the average number of eggs laid in the first year, $\bar{E}_1$, and in the second year, $\bar{E}_2$, and the regression of the average number of eggs laid in the second year on the average number laid in the first year is

$$r_{\bar{E}_1\bar{E}_2} = 0.447 \pm 0.073, \quad \bar{E}_2 = 31.147 + 0.625\ \bar{E}_1$$

The means for the two years and the regression line are represented in diagram 3.

There is, therefore, a material relationship between the average production of the flock in the second year and the average production in the first year.

The correlation between the monthly records of one year and the annual record of another year

Because of the desirability of using the records of single months as bases for the prediction of the second year we have determined the correlations between the records made by the birds in the several months of the pullet

TABLE 3

Comparison of inter-annual correlations for various months of first and second year.

MONTH OF FIRST OR OF SECOND YEAR	AVERAGE OF MONTHLY CORRELATION WHEN BIRDS ARE ARRANGED BY MONTHS OF FIRST YEAR	AVERAGE OF MONTHLY CORRELATION WHEN BIRDS ARE ARRANGED BY MONTHS OF SECOND YEAR	CORRELATION BETWEEN MONTHLY EGG RECORD OF FIRST YEAR AND ANNUAL EGG RECORD OF SECOND YEAR	CORRELATION BETWEEN MONTHLY EGG RECORD OF SECOND YEAR AND ANNUAL EGG RECORD OF FIRST YEAR
November	.0967	.1063	+.1671±.0312	+.2162±.0306
December	.1521	.0581	+.2802±.0295	+.1224±.0316
January	.1645	.1674	+.3103±.0290	+.3234±.0287
February	.1448	.1689	+.2602±.0299	+.3228±.0287
March	.1302	.1944	+.2352±.0303	+.3240±.0287
April	.1181	.1025	+.2105±.0306	+.1497±.0313
May	.1527	.1342	+.2862±.0294	+.2081±.0307
June	.1779	.1493	+.3429±.0283	+.2436±.0301
July	.1943	.1826	+.3727±.0276	+.3272±.0286
August	.2222	.2020	+.4350±.0260	+.3717±.0276
September	.1620	.2570	+.3390±.0284	+.4974±.0241
October	.1798	.1727	+.3565±.0280	+.3469±.0282

year and the record made in the whole of the second year. For the sake of completeness, and to facilitate the investigation of certain problems to

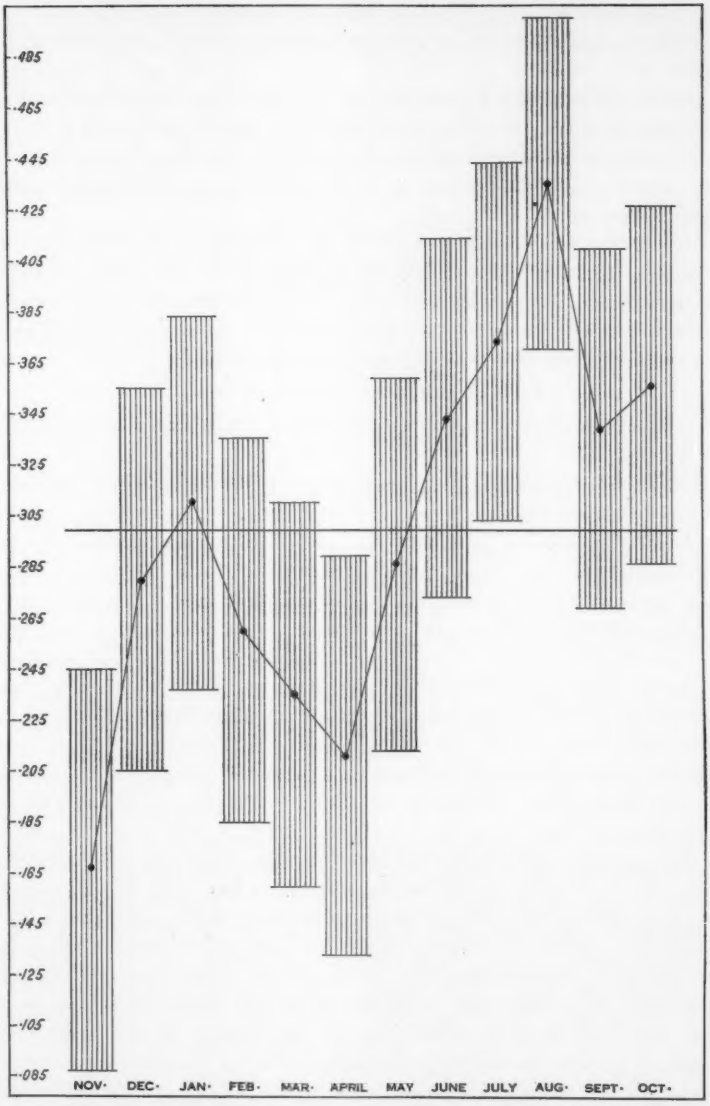

DIAGRAM 4.—Magnitudes of correlation between the production of the individual months of the first year and the annual production of the second year.

be discussed later, we have also determined the correlations between the individual months of the second year and the annual production of the first year.

The coefficients are laid beside the means of the correlations between the individual months of the year (to be discussed later) in table 3.

The results for the correlation between the individual months of the first year and the annual total of the second year are represented graphically in diagram 4.

In this diagram the magnitude of the correlations for the individual months of the first year and the annual total of the second year is shown by the relation of the solid dots to the scale of ordinates. On either side of these dots a shaded area measures a distance of two and one-half times the probable error of r as determined from the formula

$$E_r = .6745\ (1-r^2) / \sqrt{443}.$$

The heavy transverse bar represents the average value of the twelve coefficients.

The correlation for the winter months is low, increasing from November to January, after which it drops to the spring minimum in April, and then rises to the maximum values for the year in the summer months of July and August.

It appears from this diagram that the magnitude of the correlations between the individual months and the total productions of another year of the bird's life differs significantly from month to month. This conclusion is based in part on a study of the probable errors and in part on the general consistency of the results, taken alone and in comparison with those of a following section. A similar diagram has been prepared for the relationship between the individual months of the second year and the total production of the first year, but is so similar to diagram 4 that it does not merit reproduction.

The suggestion may be offered that the differences between the correlations for the records for individual months of one year and the annual totals of another year may be due to differences in the significance of the correlation coefficient. There might be a true biological relationship between the egg production of a given month of one year and the total production of another year, which would not be detected by the correlation coefficient at all if regression were non-linear. It seems desirable, therefore, to test the linearity of the regression of the annual production of one year on the production of the individual months of the other year, with the object of determining whether the differences in the magnitude of the correlations between the monthly records of one year and the annual

records of another year may be due to inadequacies in the correlation coefficient as a measure of the interrelationship in the egg-laying activities of these two periods.

Since the regression of the second-year annual record on the first-year monthly record is the one of practical significance in the prediction of future production, it alone will be considered. The regression equations are as follows:

$$E = 131.088 + .835\, e_1$$
$$E = 126.040 + 1.333\, e_2$$
$$E = 124.810 + 1.562\, e_3$$
$$E = 120.150 + 1.648\, e_4$$
$$E = 94.382 + 2.228\, e_5$$
$$E = 89.272 + 2.320\, e_6$$
$$E = 81.649 + 2.635\, e_7$$
$$E = 83.325 + 2.639\, e_8$$
$$E = 93.752 + 2.311\, e_9$$
$$E = 104.718 + 2.236\, e_{10}$$
$$E = 129.102 + 1.657\, e_{11}$$
$$E = 131.734 + 1.958\, e_{12}$$

These lines, and the mean productions of the birds for the second year for each grade of monthly egg record for the first year, are represented in diagrams 5 and 6. Certain of the means deviate very widely from the theoretical lines. These are, however, for the most part based on only a few birds, and in some cases upon only a single individual. Until further data are available, it does not seem worth while to consider regression curves of a higher order.

From the practical side it is very fortunate that the highest correlations between the production of the individual months of the first year and the annual total of the second year is found in the summer months of July and August. It is in the months of June, July and August that culling should be undertaken for the purpose of eliminating from the flock birds which will not be useful for further production during the current (pullet) year. It is at this time that the best selection of birds for the second-year's production can be made.

The practical phases of this problem of the prediction of second-year production from the records of individual months of the first year will be treated in an Experiment Station bulletin to be published shortly. The correlations between the monthly records of the first and second year will be discussed in the following section.

Inter-annual correlation for monthly production

We now have to consider the correlation between the production of the individual months of the bird's first and second year. Table 4 shows the

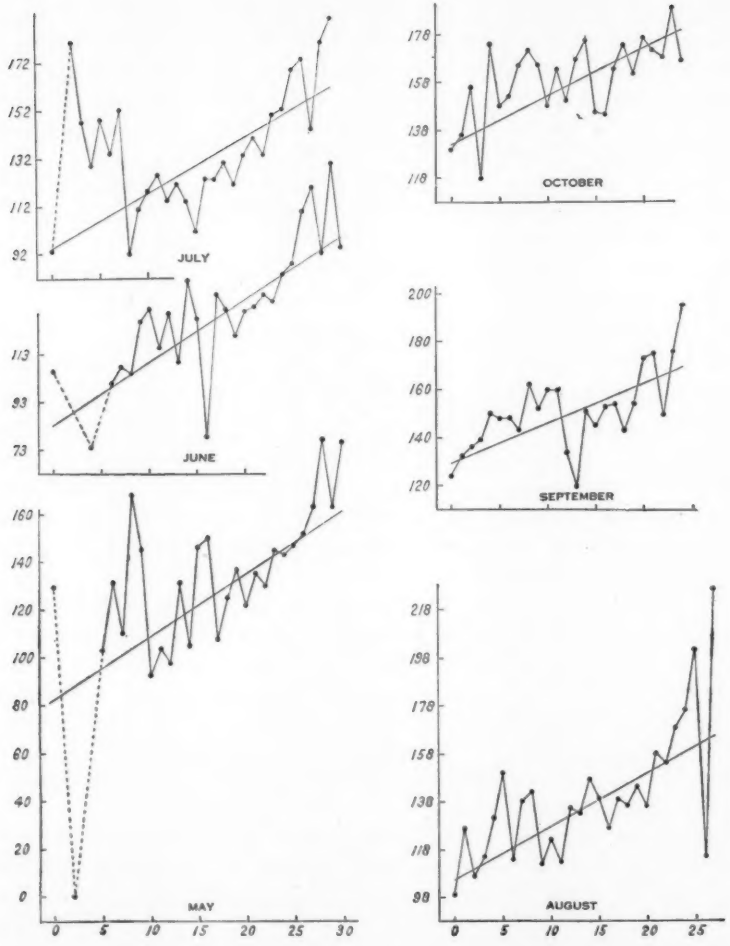

DIAGRAM 5.—Regression of second-year annual production on first-year monthly production.

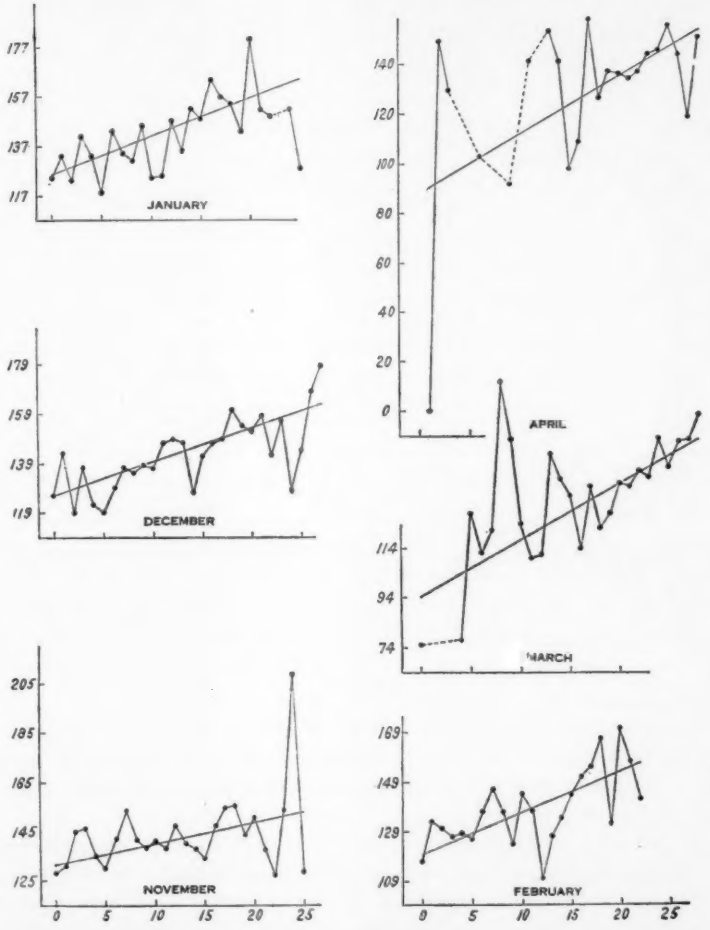

DIAGRAM 6.—Regression of second-year annual production on first-year monthly production.

TABLE 4

Correlation coefficients measuring the relationship between the egg production of the individual months of first two years of the bird's life.

Bird's first year	Bird's second year: NOVEMBER c_1	DECEMBER c_2	JANUARY c_3	FEBRUARY c_4	MARCH c_5	APRIL c_6
November	+.1100± .0317	+.0746± .0319	+.1627± .0312	+.2053± .0307	+.1686± .0311	+.0410± .0320
c_1	3.47	2.34	5.22	6.69	5.41	1.28
December	+.2017± .0307	+.1685± .0311	+.2195± .0305	+.2714± .0297	+.1504± .0313	+.0184± .0320
c_2	6.56	5.41	7.19	9.14	4.80	0.57
January	+.2031± .0307	+.1713± .0311	+.3186± .0288	+.2975± .0292	+.1656± .0312	−.0423± .0320
c_3	6.61	5.51	11.1	10.2	5.31	1.32
February	+.1605± .0312	+.2003± .0308	+.3316± .0285	+.3504± .0281	+.2373± .0302	−.0348± .0320
c_4	5.14	6.51	11.6	12.5	7.84	1.09
March	+.0744± .0319	+.0215± .0320	+.1509± .0313	+.2382± .0302	+.3348± .0285	+.0938± .0318
c_5	2.33	0.67	4.82	7.88	11.8	2.95
April	+.0325± .0320	−.0290± .0320	+.0476± .0320	+.1031± .0317	+.3175± .0288	+.2155± .0306
c_6	1.02	0.91	1.49	3.25	11.0	7.05
May	+.0166± .0320	−.0265± .0320	+.0339± .0320	+.0280± .0320	+.1922± .0309	+.2810± .0295
c_7	0.52	0.83	1.06	0.87	6.23	9.52
June	−.0177± .0320	+.0165± .0320	+.0949± .0318	+.0182± .0320	+.1842± .0310	+.2686± .0297
c_8	0.55	0.52	2.99	0.57	5.95	9.03
July	−.0444± .0320	+.0353± .0320	+.1961± .0308	+.0641± .0319	+.2147± .0306	+.2077± .0307
c_9	1.39	1.10	6.36	2.01	7.02	6.77
August	+.0814± .0318	+.0893± .0318	+.2008± .0308	+.1182± .0316	+.1661± .0312	+.1407± .0314
c_{10}	2.56	2.81	6.53	3.74	5.33	4.48
September	+.0678± .0319	−.0439± .0320	+.1265± .0315	+.1314± .0315	+.1069± .0317	+.0303± .0320
c_{11}	2.13	1.37	4.01	4.17	3.37	0.95
October	+.3895± .0272	+.0192± .0320	+.1254± .0315	+.2013± .0308	+.0940± .0318	+.0103± .0320
c_{12}	14.3	0.69	3.97	6.54	2.96	0.32

144 correlation coefficients which may be deduced as measures of this relationship.

For the purpose of determining how close the relationship between the months of the first two years of the bird's life may be, we have formed the frequency distribution of the magnitudes of the correlation coefficients shown in table 5.

These coefficients are preponderantly positive in sign. Two coefficients fall in the "zero" class, ten are negative in sign and numerically larger than −.0125, while 132 are positive in sign and larger than +.0125 numerically.

When r is very small the probable error is difficult of interpretation. We have, however, calculated the probable errors of the 144 individual coefficients by the usual formula. These probable errors are given in table 4. The ratios of the correlation coefficients to their probable errors are given (to one or two decimal places) under the individual coefficients. A glance at these entries in table 4 will show that many of the coefficients are over three times as large as their probable errors. These tests of significance (r/E_r) have been seriated according to their magnitude in table 6.

TABLE 4 (continued)

Correlation coefficients measuring the relationship between the egg production of the individual months of first two years of the bird's life.

Bird's first year	Bird's second year: MAY c_7	JUNE c_8	JULY c_9	AUGUST c_{10}	SEPTEMBER c_{11}	OCTOBER c_{12}
November	+.0765± .0319	+.0532± .0320	+.0220± .0320	+.0278± .0320	+.1256± .0315	+.0932± .0318
c_1.....	2.40	1.66	0.69	0.87	3.98	2.93
December	+.0375± .0320	+.0695± .0319	+.0889± .0318	+.1481± .0313	+.2612± .0299	+.1902± .0309
c_2.....	1.17	2.18	2.79	4.73	8.75	6.16
January	−.0384± .0320	+.0283± .0320	+.1245± .0316	+.2008± .0308	+.3208± .0288	+.2237± .0304
c_3.....	1.20	0.89	3.95	6.53	11.2	7.35
February	−.0651± .0319	+.0249± .0320	+.0989± .0317	+.1202± .0316	+.2124± .0306	+.1014± .0317
c_4.....	2.04	0.78	3.12	3.81	6.94	3.20
March	+.0789± .0319	+.0895± .0318	+.0974± .0317	+.1401± .0314	+.1610± .0312	+.0815± .0318
c_5.....	2.48	2.82	3.07	4.46	5.16	2.56
Aprtil	+.1774± .0310	+.1560± .0313	+.1062± .0317	+.1108± .0317	+.1164± .0316	+.0630± .0319
c_6...	5.71	4.99	3.35	3.50	3.68	1.97
May	+.3369± .0284	+.3191± .0288	+.2317± .0303	+.1979± .0308	+.1425± .0314	+.0796± .0318
c_7.....	11.9	11.1	7.64	6.43	4.54	2.50
June	+.3929± .0271	+.3335± .0285	+.3056± .0291	+.2810± .0295	+.2019± .0307	+.0549± .0320
c_8.....	14.5	11.7	10.5	9.52	6.57	1.72
July	+.3642± .0278	+.3869± .0273	+.3665± .0277	+.2554± .0300	+.2183± .0305	+.0673± .0319
c_9.....	13.1	14.2	13.2	8.53	7.15	2.11
August	+.2669± .0298	+.2772± .0296	+.3652± .0278	+.3410± .0283	+.3654± .0278	+.2547± .0300
c_{10}.....	8.97	9.37	13.1	12.0	13.2	8.50
September	+.0160± .0320	+.0482± .0320	+.2557± .0300	+.3354± .0284	+.4791± .0247	+.3905± .0272
c_{11}.....	0.50	1.51	8.54	11.8	19.4	14.4
October	−.0332± .0320	+.0056± .0320	+.1286± .0315	+.2651± .0298	+.4799± .0247	+.4718± .0249
c_{12}.....	1.04	0.18	4.08	8.90	19.5	18.9

This table shows that over a hundred of the positive coefficients may be considered significant in comparison with their probable errors.

Turning now to the series of correlation coefficients as a whole we may note that the statistical constants deduced from the distribution of the correlation coefficients (in table 5) are:

Mean, .1585 ± .0070
Standard deviation, .1242 ± .0049
Coefficient of variation, 78.36

It is clear, therefore, that the average value of the correlation between the individual months of the bird's first and second year is positive and deviates significantly from zero. We may confidently assert, therefore, that the relationship between the egg production of the bird's two years is so close that, within limitations to be discussed presently, any month of the first-year record furnishes some (though very slender) basis for prediction of the egg production of each individual month of the bird's second year.

It will be noted that the standard deviation of the coefficients is relatively very large, 78.36 percent of the mean. If the correlation between

all the months of the two years were biologically the same, and differences between the calculated values due merely to the probable errors of random sampling, one might expect a standard deviation of

$$\sigma'_r = \frac{1-r^2}{\sqrt{N}}$$

TABLE 5

Frequency distribution of correlations between the individual months of the bird's first and second year.

CLASS CENTERS	FREQUENCIES
−.075	1
−.050	4
−.025	5
.000	2
+.025	17
+.050	5
+.075	12
+.100	15
+.125	9
+.150	9
+.175	8
+.200	14
+.225	7
+.250	5
+.275	7
+.300	2
+.325	9
+.350	2
+.375	5
+.400	3
+.425	—
+.450	—
+.475	3

TABLE 6

Frequency distribution of ratios of correlation coefficients for individual months to their probable errors.

CLASS	FREQUENCIES
− 2.50 to 1.50	1
− 1.50 to .50	9
− .50 to .50	3
+ .50 to 1.50	18
+ 1.50 to 2.50	10
+ 2.50 to 3.50	19
+ 3.50 to 4.50	12
+ 4.50 to 5.50	12
+ 5.50 to 6.50	7
+ 6.50 to 7.50	15
+ 7.50 to 8.50	4
+ 8.50 to 9.50	8
+ 9.50 to 10.50	3
+10.50 to 11.50	5
+11.50 to 12.50	7
+12.50 to 13.50	4
+13.50 to 14.50	4
+14.50 to 15.50	—
+15.50 to 16.50	—
+16.50 to 17.50	—
+17.50 to 18.50	—
+18.50 to 19.50	3

where σ'_r is the theoretical standard deviation of r as determined on samples of $N = 443$ individuals. Taking the mean value of the correlation in the 144 permutations of months as the most probable value of r, this formula gives

$$\sigma'_r = .0463,$$

as compared with $\sigma_r = .1242$ determined directly from the constants.

It is evident that there is *differentiation* in these correlation coefficients, not merely *variations* due to the errors of random sampling. This problem

will be treated in detail in a forthcoming paper now in preparation. One aspect of the problem will be discussed in the following section.

Genetic significance of the inter-annual correlation for monthly production

We now have to consider these correlations in relation to a genetic problem, namely, that of the dependence of fertility on the presence or absence of definite Mendelian factors.

There has become current among geneticists the theory of a "winter cycle" of egg production which is often assumed to have become superimposed upon the normal "spring" or "reproductive cycle." Statements concerning the exact relations of this "cycle" to the other "cycles" of the year have left much to be desired in definiteness. A detailed review of the theory of the "winter cycle" in relation to the problem of total egg production and to the inheritance of egg production would lead us too far from our main purpose. It should be clear, however, that if birds differ in their egg production because of the presence or absence of definite Mendelian factors for winter egg production, there should be a correlation between the winter egg production of the bird's first year and that of the bird's second year. This should be true unless an additional hypothesis be adduced, namely that the genes which determine differences in the winter egg production of the pullet year have no influence on the egg production of the second year. If superiority or inferiority in laying capacity be determined primarily by these Mendelian factors for winter production, it would be reasonable to expect that the highest correlations between the production of the sub-periods of two years would be found in those periods which owed their differences to actual genotypic constitution, providing always that the second hypothesis suggested above is not adduced to explain away a result contrary to this reasonable expectation.

In short, the assumption that variation in winter egg production is largely due to zygotic composition with reference to three distinct factors which are separately inherited according to Mendelian theory seems to have certain consequences for inter-annual correlation. The egg record of the second year is distributed over a winter, spring, summer and autumn period although we have not at present evidences to indicate that the egg records of these arbitrarily delimited periods represent real biological entities. If the winter egg production of a bird's first year is determined by her zygotic formula it seems rather difficult to assume that her second period of winter production is independent of this zygotic formula. If the egg record of the winter period or cycle in both first and second year be determined by zygotic constitution, which according to assumption

may differ fundamentally from individual to individual, *one might logically expect to find an abnormally high correlation between the winter months of the entire winter periods of the two years even though the correlations for all the other months were sensibly zero.*

Our first problem is, therefore, to determine whether, as a matter of fact, there is a higher correlation between the months comprising the winter "cycles" of the first and second year than between the months of the spring, summer or autumn "cycles."

To solve this problem we have first of all determined the mean correlations between the months of the first and second year, for each month of the first year and each month of the second year, respectively. The results are shown in the first two sections of table 3. These are the averages of the twelve months of the second year associated with November, December, etc., of the first year, and of the correlations between November, December, etc., of the second year and the twelve individual months of the first year.

Our problem is to determine whether any group of these correlations can be looked upon as sensibly larger than any other group. It is evident from the tables that the averages differ somewhat among themselves. This would be expected because of the errors of random sampling, whether there be any real biological differentiation in the correlations in the individual months or not. We have, therefore, determined the standard deviation of the coefficient of correlation between the individual months of the two years from the frequency distribution shown in table 5. We have assumed for purposes of graphic representation that if the correlations for the various permutations of the months of the first and the second year were the same within the limits of the probable errors of random sampling, samples of twelve should show means with probable errors of $.6745\ \sigma_r/\sqrt{12}$.

On diagram 7 the heavy transverse lines represent the average correlation ($\bar{r} = .1585$) for the 144 combinations of first- and second-year months. The solid dots connected by lines show the values of the individual averages. The lines limiting the shaded area, above and below, show the limits of 2.5 times the probable error of the mean of samples of twelve. The light lines traversing the shaded zones are drawn to represent units of one-half the probable error.

From diagram 7 it appears that in each arrangement of the data only two averages fall outside the range of two and one-half times the probable error of the mean. In the determination of the averages according to the months of the first year there are seven determinations which fall within

the range of the probable error above and below the general mean as compared with six cases within and six cases beyond this limit which would be expected if the correlation for the individual months were random

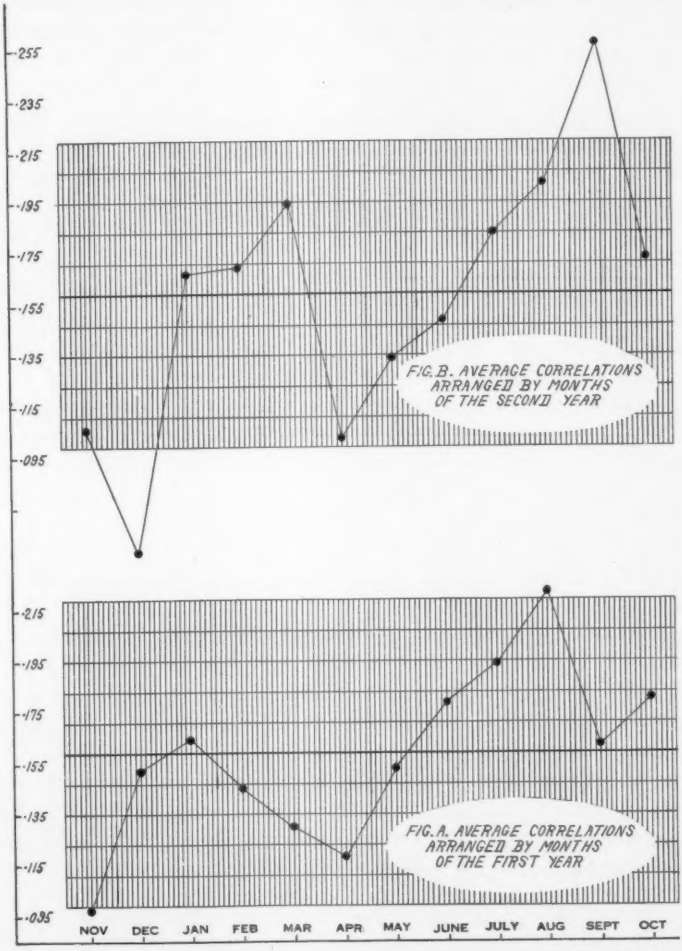

DIAGRAM 7.—Average values of inter-annual correlations for production of individual months.

samples from a series of correlations with the variability of the present series. We can not, therefore, positively assert on the basis of the probable

errors available that there is any significant difference between the average values of the inter-annual correlations of the individual months of either the first or the second year.

There is, nevertheless, at least an apparent orderly trend in the averages. In both series, November and April are conspicuously low, and the values of the averages rise from April to August or September. The average values of the inter-annual coefficients for November, December, January and February, *which, according to the Mendelian theory reviewed, should be higher than those for the other months, are as a matter of fact among the lower values*, all four coefficients for each of the two arrangements falling either below or very near the general average value of the 144 coefficients representing all the 12×12 permutations for the individual months of the first and second years.

As a further test of the differences in the correlations between the individual months, we may turn back to (a) the coefficients measuring the relationship between the production of individual months of the bird's first year and the total production of the bird's second year, and (b) the correlation between the individual months of the bird's second year and the total production of the bird's first year, as shown in the two final columns of table 3 and represented graphically in diagram 4 for the individual months of the first year.

The correlation coefficients measuring the relationship between the egg production of the individual months of the first year and the annual total of the second year show a distribution in close agreement with that found in figure A of diagram 7.

The results for the relationship between the individual months of the second year and the total production of the first year, shown in the final section of table 3 (but not represented graphically), are in excellent general agreement with the distribution of means of inter-annual correlations for individual months shown in figure B of diagram 7.

The preceding methods have the disadvantage that the entire range of months of the second-year's production associated with each month of the first year is lumped together. It is quite conceivable that the 12 general averages of correlations thus obtained do not properly represent the relation between the individual "cycles" of the first year and the individual "cycles" of the second year. It is quite conceivable that the average correlation for twelve months might give a quite erroneous measure of the correlations for the individual months. Nevertheless some average values must be considered because of the large probable errors of the constants when such fine divisions of the year are recognized.

It seems desirable, therefore, to determine the average value of the correlations between the months constituting the recognized "cycles" of production of the first year and the individual months of these "cycles" in the second year.

The periods of egg production, as classified by PEARL and SURFACE (1909) and adopted by other students of egg distribution are:

Period of winter egg production, November to February, inclusive.
Period of spring egg production, March, April and May.
Period of summer egg production, June, July and August.
Period of autumn egg production, September and October.

The averages for the correlations between the individual months of the different periods, and the standard deviations of these correlations, calculated directly from the ungrouped coefficients, appear in table 7. The values with signs are the averages of the correlations. The values without signs are the standard deviations.

TABLE 7

Average and standard deviations of inter-annual correlations for monthly periods classified according to conventional "cycles" of the first and second years.

First year	Second year: WINTER	SPRING	SUMMER	AUTUMN
Winter (November to February)	+.21545 .07656 $N=16$	+.05954 .09583 $N=12$	+.08392 .05454 $N=12$	+.19105 .07516 $N=8$
Spring (March to May)	+.05760 .07267 $N=12$	+.22534 .09324 $N=9$	+.16098 .07189 $N=9$	+.10732 .03559 $N=6$
Summer (June to August)	+.07105 .07303 $N=12$	+.24512 .08180 $N=9$	+.32359 .04336 $N=9$	+.19376 .10733 $N=6$
Autumn (September to October)	+.12713 .12184 $N=8$	+.03737 .04874 $N=6$	+.17312 .12061 $N=6$	+.45536 .03756 $N=4$

The reader who desires to do so may compare the averages for any two groups by calculating their difference in average correlation and the probable error of their difference by the usual method. Because of the relatively small number of correlations in each case we have preferred to test for differentiation between the several groups in a graphic manner.

We have worked on the assumption that, if the periods be undifferentiated, the probable error of the mean correlation for each period may be logically found from

$$E_{\bar{r}} = .6745\ \sigma_r/\sqrt{n}$$

where σ_r is the standard deviation of the 12×12 values of r and n is the number of permutations of months in the associated cycles of the two years. These numbers are shown beneath the means and standard deviations on table 7.

Remembering that for the 144 combinations of first- and second-year monthly records as deduced from the grouped correlation coefficients,

$$\bar{r} = .1585, \qquad \sigma_r = .1242$$

we have represented the mean value of the 144 constants by the position of the heavy bar on the four figures of diagram 8. On either side of the mean value we have shaded a zone limited by parallel lines above and below the general mean a distance of two and one-half times the probable error of the mean of samples of the number of correlations involved in the series of permutations in question.

Inspection of the four figures of diagram 8 shows at once that there is a distinct individuality of the birds with respect to distribution of egg production throughout the year. The figure for winter production, for example, shows that there is a closer correlation between the winter production of the first year and both the winter and autumn production of the second year than between the winter production of the first year and either the spring or summer production of the second year. The diagram for spring production shows that there is a closer correlation between the spring production of the first year and the spring production of the second year than between the spring production of the first year and the record of any other of the three formal periods of the second year, although the spring and summer periods are more closely correlated than either spring and winter or spring and autumn.

Similar results are found for the average of the correlations between the summer months of the first year and the months of the various periods of the second year. The lowest average correlation is that between the summer months of the first year and the winter months of the second year. The correlations between the summer months and the autumn months average somewhat higher than those of the summer and winter months. Finally the correlations between the summer months of the first year and the spring and summer months of the second year are the highest of the four combinations.

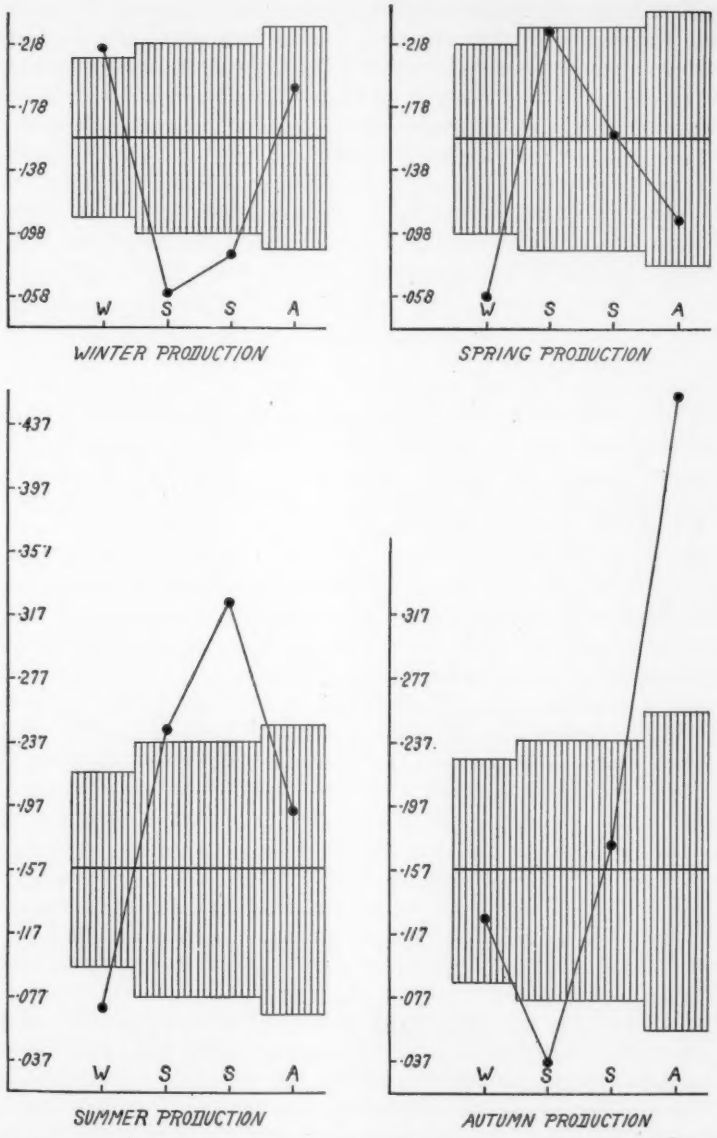

DIAGRAM 8.—Average values of inter-annual correlations for the "cycles" of the first and second year.

Comparable relationships are found when the averages of the correlations between the autumn egg records of the first year and those of the four groups of the second year are compared. The lowest average is that for autumn months and spring months. The average for autumn and summer months and for autumn and autumn months are the highest.

Summarizing the results shown by these diagrams, it is clear that (a) the average correlation for the months of the same period in the two years is in each case the highest average found in the series, that (b) the average of the correlations between the months of periods closely associated is in general higher than that of periods which are more widely separated in time, but that (c) an exception to this rule is found in the case of winter and autumn periods, which show a tendency to correlation with each other although they are the most widely separated periods. To a discussion of this point we shall return presently. For the moment it is sufficient to note that the averages show that there is a higher correlation for homologous periods of the two years. The results for the two figures of diagram 7 are therefore in some degree misleading. More detailed statistical analysis brings out relationships which are not obvious when the results are simply lumped according to the months of one year only.

Returning to the question of the significance of the deviations of these individual averages from the general average as indicated by their position with reference to the shaded area, we note that a number of the deviations would not, individually regarded, be considered statistically significant. The conclusions drawn above from the graphic tests for differentiation must rest primarily upon the consistency of results in the four diagrams. The important feature of this phase of the investigation is the conspicuous difference between the averages for the winter and spring "cycles" on the one hand and the summer and autumn "cycles" on the other. Notwithstanding the generally smaller numbers of correlations averaged in the two latter groups, the deviations from the general mean are much more conspicuous and more clearly significant. *The birds are, therefore, not primarily differentiated by winter production but by summer and autumn production.* So far as these evidences go, therefore, they suggest that in the White Leghorn heritable differences are to be looked for in the summer- and autumn-months' rather than in the winter-months' production.

We note that in this regard diagrams 7 and 8 are confirmatory. Both indicate a material increase in the closeness of correlation as the records are taken from successively later months.

The graphic method heretofore employed has the advantage that it brings out clearly the differentiation of the different periods with respect

to the intensity of correlation between the two years. To test the significance of these differences we must have recourse to a formula due to PEARSON (1906). This formula has already been suggested as a test of prepotency in inheritance studies (HARRIS 1911).

PEARSON has shown that the difference between the average of a subsample and the average of the sample from which it is drawn is given by

$$(\bar{r}_p - \bar{r}) \pm .67449 \sqrt{\frac{\sigma_r^2}{N} + \frac{\sigma_{r_p}^2}{n}\left(1 - \frac{2n}{N}\right) - \frac{n(\bar{r} - \bar{r}_p)^2}{N(N-n)}}$$

where $\bar{r}_p$ is the mean correlation in the sub-class of n individuals from the population N showing the average correlation $\bar{r}$.

The results of this test are shown in table 8.

TABLE 8

Test of significance of deviation of the average correlation for the months of the conventional cycles from the general average for all months.

PERIODS	BIRDS' FIRST AND SECOND YEAR	
	Difference and probable error $\bar{r}_p - \bar{r} \pm E_{(\bar{r}_p - \bar{r})}$	Ratio of difference to probable error $\frac{\bar{r}_p - \bar{r}}{E_{(\bar{r}_p - \bar{r})}}$
Winter—Winter	+ .0575 ± .0133	+ 4.327
Winter—Spring	− .0984 ± .0183	− 5.371
Winter—Summer	− .0740 ± .0119	− 6.240
Winter—Autumn	+ .0331 ± .0183	+ 1.813
Spring—Winter	− .1003 ± .0146	− 6.887
Spring—Spring	+ .0674 ± .0208	+ 3.243
Spring—Summer	+ .0030 ± .0166	+ 0.182
Spring—Autumn	− .0506 ± .0117	− 4.341
Summer—Winter	− .0869 ± .0147	− 5.931
Summer—Spring	+ .0872 ± .0185	+ 4.709
Summer—Summer	+ .1656 ± .0112	+14.775
Summer—Autumn	+ .0358 ± .0291	+ 1.229
Autumn—Winter	− .0308 ± .0283	− 1.091
Autumn—Spring	− .1206 ± .0145	− 8.292
Autumn—Summer	+ .0152 ± .0325	+ 0.466
Autumn—Autumn	+ .2974 ± .0139	+21.470

The entries in this table show that statistically significant deviations (> 2.5 times the probable error of the difference) of the subgroups of correlations above the average for all of the correlations, occur only in the winter-winter, spring-spring, summer-spring, summer-summer and autumn-autumn combinations. By far the highest difference, and the most clearly trustworthy difference, occurs in the autumn-autumn group.

Returning to the problem of the peculiar relationship between winter and autumn production, which has been shown by the distribution of these means to form an exception to the rule that the more closely associated periods have the higher correlation, we may turn back to the results of earlier studies on the relationship between the egg production of various periods of the year in the White Leghorn.

While the deviation of the correlation for the winter and autumn period from the average of all the periods cannot be regarded as statistically significant, it is positive in sign, i.e., the average correlation for the months of these two groups is higher than that of the 12×12 permutations of the months of the year. This reminds us of the conclusions already drawn (HARRIS, BLAKESLEE and KIRKPATRICK 1918, pp. 56-69) that there is a more intimate correlation between the egg production of the autumn and winter months (at the beginning and end of the contest year) of the same individual bird than between the egg production of these months and the production of spring and summer months.

On the basis of two-years records from the Storrs contest it has been shown (HARRIS, BLAKESLEE and KIRKPATRICK 1918, pp. 47-49, diagram 6) by the use of a special formula (HARRIS 1909) that the winter months, November, December, January and February, and the autumn months, September and October, show an increase over the theoretical quota of eggs when the annual total egg production rises above the average. The spring and summer months, March, April, May, June and July, show a lower relative contribution to the annual total than might be expected if monthly egg production were proportional to annual egg production, when the total production varies in the direction of an increase above the normal egg production of the flock as a whole. In short, the good layers owe their superiority largely to increased egg production during the autumn and winter months.

Relationship between the condition of the birds with respect to laying activity during individual months of the first year and the annual record of the second year, with special reference to culling for second-year production

The annual record of a bird depends to a very great extent upon continuous laying. In the White Leghorns at least, time of beginning of laying and time of cessation of laying are factors of great importance in determining the annual record of the birds.

In view of this fact it is desirable to determine to what extent the simple criterion of *laying* versus *not laying* in any month of the first year may be used in predicting the annual record to be expected the second year. This

method has already been applied to the investigation of first-year production (HARRIS, BLAKESLEE and KIRKPATRICK 1918, pp. 42-44, 49-56). A preliminary note (HARRIS and LEWIS 1921) has given a portion of the results of this section.

While our immediate purpose is the consideration of the second-year production of birds which did and of those which did not lay during given months of the first year, it seems desirable to give the mean for the first-year production of these birds as well. For comparison the results deduced from the data of the INTERNATIONAL EGG-LAYING CONTEST at Storrs during the years 1913-1914 and 1914-1915 (HARRIS, BLAKESLEE and KIRKPATRICK 1918, p. 42) are laid beside those presented here from the Vineland data.

TABLE 9

Mean annual and eleven-months production in first and second year for birds which did and which did not lay during individual months of the first year.

CONDITION OF BIRD IN MONTH OF FIRST YEAR	STORRS DATA FOR FIRST YEAR			VINELAND DATA FOR FIRST AND SECOND YEAR				
	Percent of flock	First-year annual mean	First-year 11-months mean	Percent of flock	First-year annual mean	First-year 11-months mean	Second-year annual mean	Second-year 11-months mean
November								
Not laying..	40.6	136.8		19.4	144.5		127.8	
Laying.....	59.4	164.2	156.2	80.6	181.2	168.2	142.7	139.9
Difference...		+27.4	+19.4		+36.7	+23.7	+14.9	+12.1
		17.9%	12.7%		21.1%	13.6%	10.7%	8.7%
December								
Not laying..	38.0	133.6		22.3	142.3		125.9	
Laying.....	62.0	165.2	155.0	77.7	183.2	169.9	143.8	141.5
Difference...		+31.6	+21.4		+40.9	+27.6	17.9+	+15.6
		20.6%	14.0%		23.9%	15.9%	12.8%	11.2%
January								
Not laying..	42.5	136.4		20.5	141.6		124.4	
Laying.....	57.5	165.6	157.2	79.5	182.4	170.4	143.8	137.3
Difference...		+29.2	+20.8		+40.8	+28.8	+19.4	+12.9
		19.1%	13.6%		23.4%	16.5%	13.9%	9.2%
February								
Not laying..	9.9	118.6		5.0	133.6		117.0	
Laying.....	90.1	157.0	152.1	95.0	176.2	163.6	141.0	131.2
Difference...		+38.4	+33.5		+42.6	+30.0	+24.0	+14.2
		25.1%	21.9%		24.5%	17.2%	17.2%	10.2%

TABLE 9 (continued)

CONDITION OF BIRD IN MONTH OF FIRST YEAR	STORRS DATA FOR FIRST YEAR			VINELAND DATA FOR FIRST AND SECOND YEAR				
	Percent of flock	First-year annual mean	First-year 11-months mean	Percent of flock	First-year annual mean	First-year 11-months mean	Second-year annual mean	Second-year 11-months mean
July								
Not laying..	2.3	72.3		3.1	110.6		92.5	
Laying.....	97.7	155.1	135.4	96.8	176.1	155.5	141.3	123.3
Difference...		+82.8	+63.1		+65.5	+44.9	+48.8	+30.8
		54.1%	41.2%		37.6%	25.8%	34.9%	22.0%
August								
Not laying..	5.1	89.9		7.2	121.6		99.5	
Laying.....	94.9	156.5	138.3	92.8	178.1	161.2	142.9	128.4
Difference...		+66.6	+48.4		+56.5	+39.6	+43.4	+28.9
		43.5%	31.6%		32.5%	22.8%	31.0%	20.7%
September								
Not laying..	23.0	115.0		33.2	147.8		124.3	
Laying.....	77.0	164.6	155.4	66.8	187.1	177.5	147.5	135.7
Difference...		+49.6	+40.4		+39.3	+29.7	+23.2	+11.4
		32.4%	26.4%		22.6%	17.1%	16.6%	8.2%
October								
Not laying..	54.7	131.9		63.2	156.9		129.8	
Laying.....	45.3	178.9	169.2	36.8	203.5	192.3	157.0	149.7
Difference...		+47.0	+37.3		+46.6	+35.4	+27.2	+19.9
		30.7%	24.3%		26.8%	20.3%	19.5%	14.2%

The essential constants are shown in table 9. This gives the percent of the flock which did and which did not lay during the months of the year in which any considerable proportion of the birds did not lay. The average annual production for these birds in the first year of both the Storrs (1913-1914 and 1914-1915) and the Vineland (1917-1918) contests and in the second year (1918-1919) of the Vineland contest are shown. The differences between these annual means for the birds of the two classes and the relative differences, i.e., the ratio of the actual differences to the annual average productions[6] of the flock are also given. While the actual differences in egg production are the data of practical significance, comparison between the three series is facilitated by expressing all of the results in comparable terms.

[6] These are 153.19 eggs for the first year at Storrs, 174.05 eggs for the first year at Vineland, and 139.79 eggs for the second year at Vineland.

The results show material differences in the annual egg record of the birds which do lay and the birds which do not lay in each of the eight months of the year which require consideration. The results for the first-year records at Storrs and the first-year records at Vineland are in fair agreement.

For the Storrs series the birds which lay in any given month show an average annual egg production from 27.4 to 82.8 eggs higher than those which do not lay, or from 17.9 to 54.1 percent. For the first year of the Vineland contest the difference in the egg production of the two groups ranges from 36.7 to 65.5 eggs or from 21.1 to 37.6 percent. Thus the difference in the annual egg production of the birds which do and which do not lay in any given month varies greatly according to the month considered. In both contests the smallest difference in the annual mean of the two groups is shown when selection is based upon the November egg record while the largest difference is that found when the segregation of the flock is based upon the July record. During the months of November, December and January the percentage differences in the annual production of the two groups of birds are higher in the Vineland than in the Storrs series, whereas for the remaining months of the eight considered, the reverse is true. The average percentage difference is 30.4 in the Storrs series and 26.6 in the Vineland series.

Turning to the second-year means we note that for each of the eight months used as a basis of selection the second-year records of birds is higher if they laid in the special month under consideration in the first year than if they did not lay in that month. The differences between the groups amount to about two dozen eggs or more in five of the eight months considered.

The fact that the difference between the second-year means for the laying and non-laying birds is always less than the difference between the first-year means may be attributed to three factors.

1. The average production is lower in the second than in the first year. This factor is obviated when the differences in the two means are expressed as relative differences.

2. Because of the imperfect correlation between the first- and second-year records, selection in the first year is not followed by as large differences in the second year as in the first year.

3. In dividing the birds into two classes,—laying and not laying,—in the first year, we necessarily reduce the number of months of oviposition from twelve to at least eleven for all of the birds of one of the classes. This classification does not, however, directly affect the second year.

To obviate the difficulty mentioned in (3) above we may determine the mean egg production for the eleven other months of the year for the months considered in table 9. The mean production of the eleven months and the mean annual production will necessarily be the same for birds which are not laying in any given month. The mean production for the eleven-month period will, however, be different and somewhat lower than the annual mean for birds which are laying in the months in question.

Consulting the table for the actual constants for the group of eleven months, we note that in the case of the first-year production at both Storrs and Vineland the differences for the birds which do and which do not lay in a given month are smaller than that for the year as a whole. This difference is shown in both the actual numbers of eggs and in the percentage differences.[7]

It seems unnecessary to consider the differences in the eleven-month totals in greater detail. All of the facts are available from table 9. The essential point which requires emphasis here is the fact that in both the Storrs and Vineland series the difference in the eleven-month totals, while not so large as those in the annual totals, are conspicuously different for birds which do and which do not lay in the specific month under consideration. *This shows clearly that the difference in the annual means represent in large part fundamental differences in the egg-laying capacity of the birds and that the observed differences in the annual means are not due primarily to the fact that one-month's egg record is necessarily fixed by the classification into the two groups.*

Turning now to the problem of the average eleven-month production of birds in the second year we find that the difference between the annual mean and the eleven-month mean is in general not so conspicuous as that observed in the case of the first-year production. This is to be expected because of the fact that in the second year the exact equivalent of the egg production of the month of the first year used as the basis of selection is not deduced from the total egg production to obtain the eleven-month average, since the production of the individual months of the first and second year are not (as shown above) perfectly correlated.

We now have to consider an interesting characteristic of these annual averages of birds which are and which are not laying in any of the later months of the year.

[7] These percentage differences are based upon the annual mean since the eleven-month average differs for each month. The basis used in determining the percent is, however, in reality too large and exaggerates somewhat the apparent difference between comparisons based upon the annual means and the eleven-month means.

For first-year's production it has been shown that the birds which are not laying in any month show a steady increase in the mean annual production as the months approach the end of the year. Similarly, and at first apparently paradoxically, the annual averages for the birds which did lay during these months also show an increase from the earlier to the later final months.

Placing all the records side by side we have the following figures:

TABLE 10

	BIRDS NOT LAYING			BIRDS LAYING		
	Storrs	Vineland I	Vineland II	Storrs	Vineland I	Vineland II
July	72.3	110.6	92.5	155.1	176.1	141.3
August	89.9	121.6	99.5	156.5	178.1	142.9
September	115.0	147.5	124.3	164.6	187.1	147.5
October	131.9	156.9	129.8	178.9	203.5	157.0

This, at first rather paradoxical result, is due to the fact that only the relatively good layers continue to lay into the later months. Thus the average of the birds which are laying is higher in August than in July and higher in September than in August and higher in October than in September. On the other hand the percentage of the flock which is not laying (table 9) increases rapidly from July and August when less than 8 percent of the birds are not laying to September when 23-33 percent of the birds are not laying and to October when 55 to 63 percent of the birds are not laying. As a result of this increase in the percentage of the birds which are not laying in these later months of the year the average quality of the birds of both classes increases rapidly with a resulting higher average annual egg production.

We may now consider as we have in a preceding publication (HARRIS, BLAKESLEE and KIRKPATRICK 1918, p. 50) the results of dividing the flock according to their egg-laying activity in more than a single month of the year. As far as records for individual months are concerned the birds may, if one desires, be divided into alternative classes,—birds which do and birds which do not lay. Birds which do lay may be logically subdivided according to the number of eggs laid. But on the basis of egg records for one month only there is no means of subdividing the class of birds which do not lay in any given month. Nevertheless, it is clear that a bird which laid during September and did not lay during October is a quite different organism from one which laid neither in September nor in October.

This problem has already been considered for the final months of the first-year data of the Storrs contest (HARRIS, BLAKESLEE and KIRKPATRICK 1918, pp. 49-53). We shall here treat both first and second years of the Vineland contest, and shall consider both the initial and the final months of the year.

Table 11 shows the laying activity in the first three months of the pullet year and the mean annual egg production of the first (pullet) and second year.

TABLE 11

First- and second-year production of birds classified according to their activity or inactivity as layers during the first three months of the pullet year.

STATUS OF BIRDS FOR FIRST YEAR	NUMBER OF BIRDS	FIRST-YEAR ANNUAL PRODUCTION	FIRST-YEAR MEAN ANNUAL PRODUCTION	SECOND-YEAR ANNUAL PRODUCTION	SECOND-YEAR MEAN ANNUAL PRODUCTION
Not laying in November.....					
Not laying in December....					
Not laying in January...	27	3020	111.9	3047	112.9
Laying in January.....	29	4329	149.3	3643	125.6
Laying in December.....					
Not laying in January...	6	816	136.0	843	140.5
Laying in January.....	24	4259	177.5	3455	144.0
Laying in November.....					
Not laying in December....					
Not laying in January...	12	1725	143.8	1456	121.3
Laying in January.....	31	5012	161.7	4322	139.4
Laying in December.....					
Not laying in January...	46	7327	159.3	5976	129.9
Laying in January.....	268	50615	188.9	39185	146.2

Disregarding two of the eight classes in which the number of individuals is too small (6 and 12, respectively) to justify conclusions concerning them, we note that the average first-year production of the six groups ranges from 111.9 eggs for the 27 birds which did not lay during the first three months of the year to 188.9 eggs for those which did lay during these three months. The values of annual egg production for other combinations of laying activity in these three months may be studied by the reader himself from the table.

It is not surprising that birds which are inactive for three months of the year should make a lower annual record during the year than birds which are laying during these months. The point of special interest is the second-year records of these birds. Such records cannot be directly influenced by activity or inactivity during the individual months of the first year.

The last column of table 11 shows that birds which did not lay during the first three months of the pullet year made an average second-year record of 112.9 eggs, as compared with an average second-year production of 146.2 eggs for the 268 birds which laid in each of these three months.

We now turn to the results for groups of two and three months at the end of the pullet year.

Table 12 shows the relationship between the laying activity in the two final months of the first year and the annual averages of first and second years.

TABLE 12

First- and second-year production of birds classified according to their activity or inactivity as layers during September and October of the pullet year.

STATUS OF BIRDS	FIRST YEAR			SECOND YEAR	
	Number of birds	Total annual production	Mean annual production	Total annual production	Mean annual production
Not laying in October.....					
Not laying in September...	131	18883	144.1	15840	120.9
Laying in September.....	149	25053	168.1	20500	137.6
Laying in October					
Not laying in September...	16	2838	177.4	2432	152.0
Laying in September.....	147	30329	206.3	23155	157.5

For the first-year records the average production increases from 144.1 eggs for birds which are inactive in both September and October to 206.3 eggs for birds which lay in both of these months. For the second year the differences are not so large, but are in full agreement with those for the first year in indicating a progressive increase in the second-year average production as the egg-laying activity of the first year is continued into the two final months of the year. Thus the average second-year production ranges from 120.9 for birds which are inactive in both September and October of the first year to 157.5 eggs for birds which lay some eggs in these final two months of the year.

Table 13 shows the relationship between the laying activity in the final three months of the first year and the average egg production of the first and second years for the 443 birds which completed the first two years of the contest. From this we note that birds which did not lay during August, September or October, had an annual egg record in the first year of 115.3 eggs. Those which laid in August but did not lay during September or October had an average of 150.6 eggs. Those which laid in August and September but did not lay in October laid 169.7 eggs in the

year. Those which laid in October and in August but not in September laid 177.4 eggs. Finally, those which laid in August, September and October laid 207.1 eggs. Two of the smaller classes, of three and five birds, respectively, are left out of account here. Thus the annual egg record of the first year depends very largely upon the laying activities in the final three months of the first year.

TABLE 13

First- and second-year production of birds classified according to their activity or inactivity as layers during August, September and October of the pullet year.

STATUS OF BIRDS FOR FIRST YEAR	NUMBER OF BIRDS	FIRST-YEAR MEAN ANNUAL PRODUCTION	SECOND-YEAR MEAN ANNUAL PRODUCTION
Not laying in October			
Not laying in September			
Not laying in August	24	115.3	91.4
Laying in August	107	150.6	127.5
Laying in September			
Not laying in August	5	122.6	102.8
Laying in August	144	169.7	138.8
Laying in October			
Not laying in September			
Not laying in August			
Laying in August	16	177.4	152.0
Laying in September			
Not laying in August	3	171.0	158.3
Laying in August	144	207.1	157.5

The second-year productions parallel very closely indeed those for the first year. Birds which do not lay in August, September or October of the first year average only 91.4 eggs during the second year, those which do not lay in September and October but do lay in August have a record of 127.5 eggs while those which lay in August and September but not in October average 138.8 eggs. Finally, birds which lay in August, September and October of the first year average 157.5 in the second year.

Thus if the birds be divided into groups in the *first year* on the basis of laying *versus* non-laying in the final three months of the year we find that the laying class will average 157.5 eggs in the second year as compared with 91.4 eggs for the non-laying class in the *second year*.

The foregoing results point clearly to an intimate relationship between the time at which egg-laying ceases in the first year and the time at which it ceases in the second year of the bird's life. This relationship is indirectly but unmistakably shown by the fact that the classification of the birds with respect to *laying activity in the initial and in the final months of the*

first influences profoundly the *record made by the groups of birds in the second year.*

Approaching the problem of the relationship between the time of the cessation of laying in the two years as directly as it can be on the basis of total production per month,[8] we may classify the birds in a two-fold manner with respect to laying activity for the initial and the final months of the two years.

The results for November egg-laying activity appear in table 14 which shows that 86 birds laid while 77 did not lay in November of the two years. Birds in which the egg-laying activity at comparable periods in the two years differs, comprise 271 which laid in November of the first year but did not lay in November of the second year, and 9 birds did not lay in November of the first year but did lay in November of the second year.

TABLE 14

November laying activities.

First year \ Second year	LAID	DID NOT LAY	TOTALS
Laid	86 +9.44	271 −9.44	357
Did not lay	9 −9.44	77 +9.44	86
Totals	95	348	443

If the November record of the second year were entirely independent of that of the first year we should expect the birds which did not lay in the second year to represent $\frac{95}{443}$ of both those which did and those which did not lay in November of the first year. Similarly, birds which did not lay in November of the second year should represent $\frac{348}{443}$ of both the group which did and the group which did not lay in the first year. The values with signs show to what extent the absolute frequencies exceed (+) or fall short of (−) expectation on the assumption that the time of beginning of laying in the two years is uncorrelated. Table 15 for December and table 16 for January are formed in the same manner.

In all cases birds which fall in the same class in the two years are more frequent than they should be on the assumption that the time of beginning of laying in the two years is uncorrelated.

[8] We shall discuss the inter-annual correlations for actual time of beginning and cessation of laying in a subsequent paper in GENETICS.

Turning now to the final months of the year we have the results for August set forth in table 17. The results for September laying activity are shown in table 18. Finally, those for October laying activity in the two years appear in table 19.

TABLE 15

December laying activities

First year	Second year: LAID	Second year: DID NOT LAY	TOTALS
Laid	98 +11.81	246 −11.81	344
Did not lay	13 −11.81	86 +11.81	99
Totals	111	332	443

TABLE 16

January laying activities.

First year	Second year: LAID	Second year: DID NOT LAY	TOTALS
Laid	279 +22.35	73 −22.35	352
Did not lay	44 −22.35	47 +22.35	91
Totals	323	120	443

TABLE 17

August laying activities.

First year	Second year: LAID	Second year: DID NOT LAY	TOTALS
Laid	371 +10.10	40 −10.10	411
Did not lay	18 −10.10	14 +10.10	32
Totals	389	54	443

In all these tables the diagonal frequencies, representing similar conditions in the two years, are far in excess of expectation. Applying the χ^2 test (PEARSON 1900) for goodness of fit, we find

November, $\chi^2 = 7.636$, P = .005435
December, $\chi^2 = 9.655$, P = .022267
January, $\chi^2 = 34.978$, P = < .000001
August, $\chi^2 = 32.097$, P = .000001
September, $\chi^2 = 60.859$, P = .000001
October, $\chi^2 = 69.628$, P = .000001

Here the values of P are taken directly from ELDERTON'S (1901) table.

TABLE 18

September laying activities.

First year	Second year: LAID	Second year: DID NOT LAY	TOTALS
Laid	239 +35.88	57 −35.88	296
Did not lay	65 −35.88	82 +35.88	147
Totals	304	139	443

TABLE 19

October laying activities.

First year	Second year: LAID	Second year: DID NOT LAY	TOTALS
Laid	90 +39.22	73 −39.22	163
Did not lay	48 −39.22	232 +39.22	280
Totals	138	305	443

We note that values of χ^2 increase from November to January and from August to October, thus indicating that the relationship between the laying activity of the two years becomes more intimate as the season progresses at the beginning of the year, and as it draws to a close. The value of P, the probability that deviation systems as great as or greater than those actually observed might have arisen through the errors of random sampling, are so small that it is impossible to consider the status of the birds in the two years as independent. The time of beginning of laying and of cessation of laying in the two years is evidently correlated.

DISCUSSION AND CONCLUSIONS

The purpose of this paper has been to present the results of an investigation of the relationship between the first- and second-year's egg production

of the domestic fowl. Published data are reviewed and the results for a flock of 443 White Leghorn fowls completing the first two years of the VINELAND INTERNATIONAL EGG-LAYING AND BREEDING CONTEST are analyzed.

All available series of data comprising reasonably large numbers of birds show a positive correlation between the total egg production of the first and second year. In the present series this interrelationship is measured for the contest flock as a whole by a coefficient of correlation of $r = .539$ to $r = .554$ and by a correlation ratio of $\eta = .559$ to $\eta = .582$, depending on the method of grouping adopted. The conclusions of those who have maintained that there is no sensible relationship between the egg record of the first and second year are, therefore, groundless.

The relationship between the first- and second-year's annual record may be fairly satisfactorily expressed by a straight-line equation.

The correlation within the individual pens is highly variable. This is readily explained as a result of the small number of birds per pen. The coefficients are, however, predominantly positive in sign and on an average slightly larger than that demonstrated for the flock as a whole. It is, therefore, clear that inter-annual correlation for the flock as a whole is due to the individuality of the birds within the pens and not to differentiation of the constituent pens of which the contest flock is composed.

Among poultrymen the opinion has long obtained that birds which lay poorly during the first year may lay more heavily during their second year, or, conversely, that birds which have laid unusually heavily during the first year will make a poor record the second year. *A priori* there might seem to be good morphological and physiological grounds for this belief. It might easily be assumed that unusually heavy laying in the first year would so reduce the number of ovules available for extrusion during the earlier periods of life that fewer would be available for laying the second year. If the problem be considered from the physiological side, it might be assumed that unusually heavy first-year laying would so tax the system of the organism that a heavy second-year record would be impossible. Again, much must depend upon the time of hatching and the time of beginning of laying. If the laying period of the first year be much shortened because of late hatching, it would seem reasonable to suppose that the egg record of the first year might also be unusually small while that of the second year might be unusually large, because of the continuation into the second year of egg-laying activity which normally would have ceased with the autumn of the first year. The results *of the present investigation* show, however, that the mean productions of the first and second years

of the individual flocks are correlated. Thus in so far as those contest birds may be considered typical of flocks in general, it appears that birds drawn from a flock of higher first-year's performance may be expected in general to give good results the second year. This may not be universally true, however, and the whole problem of the relationship between first- and second-year flock means deserves further investigation.

The correlations between the monthly productions of the first year and the annual productions of the second year are positive and statistically significant throughout. Numerically the values range from $r = +.167$ to $r = +.435$. The higher values are found in the months of June to October.

The regression of the total production of the second year on the monthly production of the first year is, roughly speaking, linear. It is, therefore, quite feasible to use the record of any individual month of the first year for predicting the annual production of the second year.

The 144 correlation coefficients measuring the relationship between the 12 individual months of the first year and the 12 individual months of the second year have been determined. These are numerically low, but with few exceptions are positive in sign. These correlations have the average value of $\bar{r} = .1585 \pm .0070$. These results show how slender is the support for the assumptions made by geneticists, of the absence of correlation between the production of the first and second laying year. The evidence indicates that the production of every individual month of the second year is correlated with that of every individual month of the first year.

The measures of interrelationship between the several months of the first and second year are not, however, of equal magnitude. This is shown by the fact that, if the true value of the 144 correlation coefficients were biologically the same and the differences in the empirically determined constants due to the errors of random sampling only, we should have a theoretical standard deviation of r of $\sigma_r' = 1 - r^2/\sqrt{443}$. Assuming that the true value of r is given to a very close approximation by the average of the 144 empirically determined coefficients, we find $\sigma_r' = 0.0463$, whereas the value computed from the actual coefficients is $\sigma_r = 0.1242$. Since the assumed true value of r is near the lower end of its possible range, it is clear that other assumptions as to the true value of the inter-annual correlation for periods of one-month duration will on the average give an even greater discrepancy between observation and expectation.

It is, therefore, clear that there are fundamental biological differences in the correlations due to the various permutations of months of the first and the second year of the bird's life.

These differences will be considered in greater detail in a subsequent paper, now in an advanced stage of preparation. In this place, we shall limit our discussion of the point to a consideration of the correlations for the conventional cycles or periods heretofore recognized by geneticists. Grouping the constants according to the conventional winter, spring, summer and autumn "cycles" or periods we find the following deviations of the averages of the correlations between the months of the several periods from the average for the series of 144 correlations (table 20).

TABLE 20

First year \ Second year	WINTER	SPRING	SUMMER	AUTUMN
Winter	+.0575	−.0984	−.0740	+.0331
Spring	−.1003	+.0674	+.0030	−.0506
Summer	−.0869	+.0872	+.1656	+.0358
Autumn	−.0308	−.1206	+.0152	+.2974

The deviations which may be considered statistically significant (greater than three times the probable error) are printed in heavy type. It is clear that the egg records of homologous periods of the year are more closely correlated than are those of non-homologous periods. The constants show, however, that in White Leghorn fowl it is the autumn rather than the winter periods of the two years which are most highly correlated.

The existence of a correlation between the production of the first and of the second year of the bird's life makes possible the prediction of the average second-year record of birds whose first-year production is known. The economic application of the principle of inter-annual correlation to the culling of the flock for the purpose of increasing second-year production will be discussed elsewhere. In this paper it has been shown that the relationship between the first- and the second-year laying activity is such that the mere classification of the birds into the alternative groups of "laying" and "not laying" in a number of the months of the first year may serve for the separation of groups of birds which will differ materially in their second-year production.

Even greater differences may be obtained in the second year when the birds are grouped according to their activity or non-activity as layers in more than a single month of the first year.

Division into alternative classes with respect to laying activity in the two years indicates that there is a correlation between both the time of beginning and the time of cessation of laying in the first and second year.

A more exhaustive investigation of a number of the problems considered in this paper is now in progress.

LITERATURE CITED

BALL, E. D., and ALDER, B., 1917 Breeding for egg production. Part II. Seasonal distribution of egg production with especial reference to "winter" egg production. Bull. Utah Agric. Coll. Exp. Sta. 149, 71 pp.

BALL, E. D., ALDER, B., and EGBERT, A. D., 1916 Breeding for egg production. Part I. A study of annual and total production. Bull. Utah Agric. Coll. Exp. Sta. 148, 60 pp.

BALL, E. D., TURPIN, G., and ALDER, B., 1914 A study in annual egg production. Bull. Utah Agric. Coll. Exp. Sta. 135, 44 pp.

BLAKEMAN, J., 1905 On tests for linearity of regression in frequency distributions. Biometrika **4**: 332–350.

BLAKESLEE, A. F., HARRIS, J. ARTHUR, WARNER, D. E., and KIRKPATRICK, W. F., 1917 Pigmentation and other criteria for the selection of laying hens. Bull. Storrs Agric. Exp. Sta. 92, pp. 95–194.

ELDERTON, W. P., 1901 Tables for testing the goodness of fit of theory to observation. Biometrika **1**: 155–163. Republished in PEARSON'S Tables for statisticians and biometricians.

GIBSON, W., 1906 Tables for facilitating the computation of probable errors. Biometrika **4**: 385–393; also in PEARSON'S Tables for statisticians and biometricians.

GOWELL, G. M., 1903 Breeding for egg production. Bull. Maine Agric. Exp. Sta. 93, pp. 69–76.

HARRIS, J. ARTHUR, 1909 The correlation between a variable and the deviation of a dependent variable from its probable value. Biometrika **6**: 438-443.

1911 A coefficient of individual prepotency for students of heredity. Amer. Nat. **45**: 471–478.

1913 On the calculation of intra-class and inter-class coefficients of correlation from class moments when the number of possible combinations is large. Biometrika **9**: 446–472.

HARRIS, J. ARTHUR, BLAKESLEE, A. F., and WARNER, D. E., 1917 Body pigmentation and egg production in the fowl. Proc. Nation. Acad. Sci. **3**: 237–241.

HARRIS, J. ARTHUR, BLAKESLEE, A. F., WARNER, D. E., and KIRKPATRICK, W. F., 1917 The correlation between body pigmentation and egg production in the domestic fowl. Genetics **2**: 36-77.

HARRIS, J. ARTHUR, BLAKESLEE, A. F., and KIRKPATRICK, W. F., 1918 The correlation between egg production during various periods of the year in the domestic fowl. Genetics **3**: 27-72.

HARRIS, J. ARTHUR, KIRKPATRICK, W. F., BLAKESLEE, A. F., WARNER, D. E., and CARD, L. E., 1921 The egg records of limited periods as criteria for predicting the egg production of the White Leghorn fowl. Genetics **6**: 265–309.

HARRIS, J. ARTHUR, KIRKPATRICK, W. F., and BLAKESLEE, A. F., 1921 The prediction of annual egg production from the records of limited periods. Proc. Nation. Acad. Sci. **7**: 213–219.

HARRIS, J. ARTHUR, and LEWIS, H. R., 1921 The second-year record of birds which did and which did not lay during individual months of the pullet year. Science N.S. **54**: 224–226.

LEWIS, HARRY R., 1917 The Vineland International Egg Laying and Breeding Contest, (Rules and Regulations). Contest Series, New Jersey State Agric. Exp. Sta. Bull. 1.

1918 The Vineland International Egg Laying and Breeding Contest. Hints to Poultrymen, Volume 5, Number 2. New Jersey State Agric. Exp. Sta.

LEWIS, HARRY R., HANNAS, RALSTON R., and WENE, ELMER H., 1920 The first two years of the Vineland contest. New Jersey Agric. Exp. Sta. Bull. 338.

NIXON, CLARA, 1912 A study of the first, second and third year egg production of White Leghorn hens. Ann. Rep. Amer. Breed. Assoc. 7: 279–288.

PEARL, R., 1912 The mode of inheritance of fecundity in the domestic fowl. Jour. Exp. Zoöl. 13: 153–268.

PEARL, R., and SURFACE, F. M., 1909 A biometrical study of egg production in the domestic fowl. I. Variation in annual egg production. Bull. Bur. Animal Indust., U. S. Dept. Agric. 110, pt. 1, 1–80.

PEARSON, K. 1900 On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil. Mag. 50: 157–175.

1905 Mathematical contributions to the theory of evolution. On the general theory of skew correlation and non-linear regression. Drapers' Co., Research Mem., Biometric Ser., 2, pp. 1–54, diagrams 1–5, London.

1906 Note on the significant or non-significant character of a sub-sample drawn from a sample. Biometrika 5: 181–183.